frontiers | Research Topics

Big data analytics for smart healthcare applications, 2nd edition

Edited by
Celestine Iwendi, Thippa Reddy Gadekallu and Ali Kashif Bashir

Published in
Frontiers in Public Health
Frontiers in Computer Science
Frontiers in Oncology

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-7275-7
DOI 10.3389/978-2-8325-7275-7

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Big data analytics for smart healthcare applications, 2nd edition

Topic editors

Celestine Iwendi — School of Creative Technologies University of Bolton, United Kingdom

Thippa Reddy Gadekallu — Department of Electrical and Computer Engineering, Lebanese American University, Lebanon

Ali Kashif Bashir — Manchester Metropolitan University, United Kingdom

Citation

Iwendi, C., Gadekallu, T. R., Bashir, A. K., eds. (2025). *Big data analytics for smart healthcare applications, 2nd edition*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7275-7

Publisher's note: This is a 2nd edition due to an article retraction.

Table of contents

14 Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate
Jayakumar Kaliappan, Kathiravan Srinivasan, Saeed Mian Qaisar, Karpagam Sundararajan, Chuan-Yu Chang and Suganthan C

26 Blockchain-Secured Recommender System for Special Need Patients Using Deep Learning
Eric Appiah Mantey, Conghua Zhou, Joseph Henry Anajemba, Izuchukwu M. Okpalaoguchi and Onyeachonam Dominic-Mario Chiadika

38 Recurrent Neural Network and Reinforcement Learning Model for COVID-19 Prediction
R. Lakshmana Kumar, Firoz Khan, Sadia Din, Shahab S. Band, Amir Mosavi and Ebuka Ibeke

50 A Predictive and Preventive Model for Onset of Alzheimer's Disease
Udit Singhania, Balakrushna Tripathy, Mohammad Kamrul Hasan, Noble C. Anumbe, Dabiah Alboaneen, Fatima Rayan Awad Ahmed, Thowiba E. Ahmed and Manasik M. Mohamed Nour

62 Fischer Linear Discrimination and Quadratic Discrimination Analysis–Based Data Mining Technique for Internet of Things Framework for Healthcare
Mohammad Kamrul Hasan, Taher M. Ghazal, Ali Alkhalifah, Khairul Azmi Abu Bakar, Alireza Omidvar, Nazmus S. Nafi and Johnson I. Agbinya

80 A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients
Dong Wang, Jinbo Li, Yali Sun, Xianfei Ding, Xiaojuan Zhang, Shaohua Liu, Bing Han, Haixu Wang, Xiaoguang Duan and Tongwen Sun

88 Smart Cardiac Framework for an Early Detection of Cardiac Arrest Condition and Risk
Apeksha Shah, Swati Ahirrao, Sharnil Pandya, Ketan Kotecha and Suresh Rathod

106 Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic
Ram Krishn Mishra, Siddhaling Urolagin, J. Angel Arul Jothi, Ashwin Sanjay Neogi and Nishad Nawaz

120 Forecasting Dengue Hotspots Associated With Variation in Meteorological Parameters Using Regression and Time Series Models
Seema Patil and Sharnil Pandya

136 COVIDSAVIOR: A Novel Sensor-Fusion and Deep Learning Based Framework for Virus Outbreaks
Sharnil Pandya, Anirban Sur and Nitin Solke

149 A Comprehensive Review on the Optical Micro-Electromechanical Sensors for the Biomedical Application
Anup M. Upadhyaya, Mohammad Kamrul Hasan, S. Abdel-Khalek, Rosilah Hassan, Maneesh C. Srivastava, Preeta Sharan, Shayla Islam, Asma Mohammed Elbashir Saad and Nguyen Vo

177 NextGen Public Health Surveillance and the Internet of Things (IoT)
Kirti Sundar Sahu, Shannon E. Majowicz, Joel A. Dubin and Plinio Pelegrini Morita

186 Trustworthy Intrusion Detection in E-Healthcare Systems
Faiza Akram, Dongsheng Liu, Peibiao Zhao, Natalia Kryvinska, Sidra Abbas and Muhammad Rizwan

196 Neural Network Based Mental Depression Identification and Sentiments Classification Technique From Speech Signals: A COVID-19 Focused Pandemic Study
Syed Thouheed Ahmed, Dollar Konjengbam Singh, Syed Muzamil Basha, Emad Abouel Nasr, Ali K. Kamrani and Mohamed K. Aboudaif

205 Ontology-Enabled Emotional Sentiment Analysis on COVID-19 Pandemic-Related Twitter Streams
Senthil Kumar Narayanasamy, Kathiravan Srinivasan, Saeed Mian Qaisar and Chuan-Yu Chang

220 Machine Learning Applications for the Prediction of Bone Cement Leakage in Percutaneous Vertebroplasty
Wenle Li, Jiaming Wang, Wencai Liu, Chan Xu, Wanying Li, Kai Zhang, Shibin Su, Rong Li, Zhaohui Hu, Qiang Liu, Ruogu Lu and Chengliang Yin

228 An Explainable Intelligence Driven Query Prioritization Using Balanced Decision Tree Approach for Multi-Level Psychological Disorders Assessment
Sushruta Mishra, Hrudaya Kumar Tripathy, Hiren Kumar Thakkar, Deepak Garg, Ketan Kotecha and Sharnil Pandya

254 Development and Internal Validation of a Nomogram to Predict Mortality During the ICU Stay of Thoracic Fracture Patients Without Neurological Compromise: An Analysis of the MIMIC-III Clinical Database
Haosheng Wang, Yangyang Ou, Tingting Fan, Jianwu Zhao, Mingyang Kang, Rongpeng Dong and Yang Qu

265 Hybrid Deep Neural Network for Handling Data Imbalance in Precursor MicroRNA
Elakkiya R., Deepak Kumar Jain, Ketan Kotecha, Sharnil Pandya, Sai Siddhartha Reddy, Rajalakshmi E., Vijayakumar Varadarajan, Aniket Mahanti and Subramaniyaswamy V

277 Machine Learning Assisted Cervical Cancer Detection
Mavra Mehmood, Muhammad Rizwan, Michal Gregus ml and Sidra Abbas

291 Detecting COVID-19-Related Fake News Using Feature Extraction
Suleman Khan, Saqib Hakak, N. Deepa, B. Prabadevi, Kapal Dev and Silvia Trelova

300 Analysis of Tracheobronchial Diverticula Based on Semantic Segmentation of CT Images *via* the Dual-Channel Attention Network
Maoyi Zhang, Changqing Ding and Shuli Guo

313 The Construction and Development of a Clinical Prediction Model to Assess Lymph Node Metastases in Osteosarcoma
Wenle Li, Shengtao Dong, Bing Wang, Haosheng Wang, Chan Xu, Kai Zhang, Wanying Li, Zhaohui Hu, Xiaoping Li, Qiang Liu, Rilige Wu and Chengliang Yin

323 A Web-Based Prediction Model for Overall Survival of Elderly Patients With Malignant Bone Tumors: A Population-Based Study
Jie Tang, JinKui Wang and Xiudan Pan

335 Global Trends and Hotspots in Research on Rehabilitation Robots: A Bibliometric Analysis From 2010 to 2020
Xiali Xue, Xinwei Yang, Zhongyi Deng, Huan Tu, Dezhi Kong, Ning Li and Fan Xu

352 A Hybrid Framework for Intrusion Detection in Healthcare Systems Using Deep Learning
M. Akshay Kumaar, Duraimurugan Samiayya, P. M. Durai Raj Vincent, Kathiravan Srinivasan, Chuan-Yu Chang and Harish Ganesh

370 An Artificial Intelligence Approach Toward Food Spoilage Detection and Analysis
Ekta Sonwani, Urvashi Bansal, Roobaea Alroobaea, Abdullah M. Baqasah and Mustapha Hedabou

383 A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Hypopharyngeal Squamous Cell Carcinomas: A Population-Based Study
JinKui Wang, XiaoZhu Liu, Jie Tang, Qingquan Zhang and Yuanyang Zhao

394 COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques
Zunera Jalil, Ahmed Abbasi, Abdul Rehman Javed, Muhammad Badruddin Khan, Mozaherul Hoque Abul Hasanat, Khalid Mahmood Malik and Abdul Khader Jilani Saudagar

408 A Recommendation System Based on AI for Storing Block Data in the Electronic Health Repository
Vinodhini Mani, C. Kavitha, Shahab S. Band, Amir Mosavi, Paul Hollins and Selvashankar Palanisamy

420 An Empirical Model to Predict the Diabetic Positive Using Stacked Ensemble Approach
Sivashankari R., Sudha M., Mohammad Kamrul Hasan, Rashid A. Saeed, Suliman A. Alsuhibany and Sayed Abdel-Khalek

433 An Artificial Intelligence Approach for Expurgating Edible and Non-Edible Items
Dilip Kumar, Urvashi Bansal, Roobaea S. Alroobaea, Abdullah M. Baqasah and Mustapha Hedabou

446 Development and Validation of a Nomogram to Predict Distant Metastasis in Elderly Patients With Renal Cell Carcinoma
Jinkui Wang, Chenghao Zhanghuang, Xiaojun Tan, Tao Mi, Jiayan Liu, Liming Jin, Mujie Li, Zhaoxia Zhang and Dawei He

458 A Nomogram for Predicting Cancer-Specific Survival of Osteosarcoma and Ewing's Sarcoma in Children: A SEER Database Analysis
Jinkui Wang, Chenghao Zhanghuang, Xiaojun Tan, Tao Mi, Jiayan Liu, Liming Jin, Mujie Li, Zhaoxia Zhang and Dawei He

470 Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network
Ahila A, Poongodi M, Mounir Hamdi, Sami Bourouis, Kulhanek Rastislav and Faizaan Mohmed

480 Energy Expenditure Estimation of Tabata by Combining Acceleration and Heart Rate
Yiping Yan and Qingguo Chen

493 A Novel Extra Tree Ensemble Optimized DL Framework (ETEODL) for Early Detection of Diabetes
Monika Arya, Hanumat Sastry G, Anand Motwani, Sunil Kumar and Atef Zaguia

506 Development and Validation of a Competitive Risk Model in Elderly Patients With Chromophobe Cell Renal Carcinoma: A Population-Based Study
Jinkui Wang, Chenghao Zhanghuang, Xiaojun Tan, Tao Mi, Jiayan Liu, Liming Jin, Mujie Li, Zhaoxia Zhang and Dawei He

516 Development and Validation of a Nomogram to Predict Cancer-Specific Survival for Middle-Aged Patients With Early-Stage Hepatocellular Carcinoma
Chong Wen, Jie Tang and Hao Luo

526 Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models
C. Kavitha, Vinodhini Mani, S. R. Srividhya, Osamah Ibrahim Khalaf and Carlos Andrés Tavera Romero

539 A Multi-Agent Formalism Based on Contextual Defeasible Logic for Healthcare Systems
Salwa Muhammad Akhtar, Makia Nazir, Kiran Saleem, Rana Zeeshan Ahmad, Abdul Rehman Javed, Shahab S. Band and Amir Mosavi

553 COVID-19 Identification System Using Transfer Learning Technique With Mobile-NetV2 and Chest X-Ray Images
Mahmoud Ragab, Samah Alshehri, Gamil Abdel Azim, Hibah M. Aldawsari, Adeeb Noor, Jaber Alyami and S. Abdel-khalek

568 Evaluation of the Predictors for Unfavorable Clinical Outcomes of Degenerative Lumbar Spondylolisthesis After Lumbar Interbody Fusion Using Machine Learning
Shengtao Dong, Yinghui Zhu, Hua Yang, Ningyu Tang, Guangyi Huang, Jie Li and Kang Tian

579 Improving Alzheimer's Disease Detection for Speech Based on Feature Purification Network
Ning Liu, Zhenming Yuan and Qingfeng Tang

588 A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Clear Cell Renal Cell Carcinoma: A Population-Based Study
Chenghao Zhanghuang, Jinkui Wang, Zhaoxia Zhang, Liming Jin, Xiaojun Tan, Tao Mi, Jiayan Liu, Mujie Li and Dawei He

599 Cognitive Impairment of Patient With Neurological Cerebrovascular Disease Using the Artificial Intelligence Technology Guided by MRI
Lifang Zhang, Yanran Li, Lin Bian, Qingrong Luo, Xiaoxi Zhang and Bing Zhao

613 A Decision Support System for Diagnosing Diabetes Using Deep Neural Network
Osama Rabie, Daniyal Alghazzawi, Junaid Asghar, Furqan Khan Saddozai and Muhammad Zubair Asghar

626 Retinal Vessel Extraction *via* Assisted Multi-Channel Feature Map and U-Net
Surbhi Bhatia, Shadab Alam, Mohammed Shuaib, Mohammed Hameed Alhameed, Fathe Jeribi and Razan Ibrahim Alsuwailem

639 Classification of Non-Functional Requirements From IoT Oriented Healthcare Requirement Document
Iqra Khurshid, Salma Imtiaz, Wadii Boulila, Zahid Khan, Almas Abbasi, Abdul Rehman Javed and Zunera Jalil

653 Analysis of Factors Affecting Cranial Nerve Function of Patients With Vascular Mild Cognitive Impairment Through Functional Magnetic Resonance Imaging Under Artificial Intelligence Environment
Lifang Zhang, Yanran Li, Lin Bian, Qingrong Luo, Xiaoxi Zhang and Bing Zhao

662 A Study of Family Process Factors of Social Anxiety on the Internet Based on Big Data—Take Guangxi University Students as an Example
Dexu Bi

669 The Text Mining Technique Applied to the Analysis of Health Interventions to Combat Congenital Syphilis in Brazil: The Case of the "Syphilis No!" Project
Marcella A. da Rocha, Marquiony M. dos Santos, Raphael S. Fontes, Andréa S. P. de Melo, Aliete Cunha-Oliveira, Angélica E. Miranda, Carlos A. P. de Oliveira, Hugo Gonçalo Oliveira, Cristine M. G. Gusmão, Thaísa G. F. M. S. Lima, Rafael Pinto, Daniele M. S. Barros and Ricardo A. de M. Valentim

688 An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction
Junaid Rashid, Saba Batool, Jungeun Kim, Muhammad Wasif Nisar, Amir Hussain, Sapna Juneja and Riti Kushwaha

708 A Classification Algorithm-Based Hybrid Diabetes Prediction Model
Michael Onyema Edeh, Osamah Ibrahim Khalaf, Carlos Andrés Tavera, Sofiane Tayeb, Samir Ghouali, Ghaida Muttashar Abdulsahib, Nneka Ernestina Richard-Nnabu and AbdRahmane Louni

715 Identification of a Four-Gene Metabolic Signature to Evaluate the Prognosis of Colon Adenocarcinoma Patients
Yang Zheng, Rilige Wu, Ximo Wang and Chengliang Yin

728 Deep Data Analysis-Based Agricultural Products Management for Smart Public Healthcare
Wenjing Yan, Zesheng Zhang, Qingchuan Zhang, Ganggang Zhang, Qiaozhi Hua and Qiao Li

737 Medical Image Segmentation Algorithm for Three-Dimensional Multimodal Using Deep Reinforcement Learning and Big Data Analytics
Weiwei Gao, Xiaofeng Li, Yanwei Wang and Yingjie Cai

746 Decision Support System for Predicting Survivability of Hepatitis Patients
Fahad R. Albogamy, Junaid Asghar, Fazli Subhan, Muhammad Zubair Asghar, Mabrook S. Al-Rakhami, Aurangzeb Khan, Haidawati Mohamad Nasir, Mohd Khairil Rahmat, Muhammad Mansoor Alam, Adidah Lajis and Mazliham Mohd Su'ud

761 Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection
Shivaji D. Pawar, Kamal K. Sharma, Suhas G. Sapate, Geetanjali Y. Yadav, Roobaea Alroobaea, Sabah M. Alzahrani and Mustapha Hedabou

774 Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease
Michael Onyema Edeh, Surjeet Dalal, Imed Ben Dhaou, Charles Chuka Agubosim, Chukwudum Collins Umoke, Nneka Ernestina Richard-Nnabu and Neeraj Dahiya

787 Testing the Stochastic Convergence of Ecological Indicators in BRICS in the Perspective of Public Health
Jinshun Wu

809 Application of Smart Healthcare in Comparative Analysis of Effect of Early External Fixator and Plate Internal Fixation Treatment on Postoperative Complications and Lower Limb Function Recovery of Patients With Unstable Pelvic Fracture
Hongwei Zhang, Fang Sun and Yao Li

817 Analyzing the Effect of People Utilizing Mobile Technology to Make Banking Services More Accessible
Jiale Zhu and Manyi Wang

826 The Mediating Role of New Media Engagement in This Digital Age
Zhang Yujie, Megat Al Imran Yasin, Syed Agil Bin Shekh Alsagoff and Ang Lay Hoon

835 Financial Budgets of Technology-Based SMEs From the Perspective of Sustainability and Big Data
Guobiao Zhang and Tao Wang

844 Integrated Blockchain-Deep Learning Approach for Analyzing the Electronic Health Records Recommender System
Eric Appiah Mantey, Conghua Zhou, S. R. Srividhya, Sanjiv Kumar Jain and B. Sundaravadivazhagan

854 Effect of Network-Based Positive Psychological Nursing Model Combined With Elemene Injection on Negative Emotions, Immune Function and Quality of Life in Lung Cancer Patients Undergoing Chemotherapy in the Era of Big Data
Shilu Yang, Lijuan Zheng, Yan Sun and Zhuoyun Li

861 Research on Diagnosis and Management of Postgraduates Mental Health Status Based on BP Neural Network
Jiangning Xie

869 Social Intervention and Governance of Youth School Bullying—Based on Computer Medical Data Analysis
Jiahui Zhao

881 The Relationship Between Adjustment and Mental Health of Chinese Freshmen: The Mediating Effect of Security and the Moderating Effect of Gender
Liu Cao

887 Deep Conviction Systems for Biomedical Applications Using Intuiting Procedures With Cross Point Approach
Hariprasath Manoharan, Shitharth Selvarajan, Ayman Yafoz, Hassan A. Alterazi, Mueen Uddin, Chin-Ling Chen and Chih-Ming Wu

901 A Visualized Dynamic Prediction Model for Overall Survival in Elderly Patients With Pancreatic Cancer for Smart Medical Services
Jiang Zhong, XingShu Liao, Shuang Peng, Junyi Cao, Yue Liu, Chunyang Liu, Ju Qiu, Xiaoyan Guan, Yang Zhang, Xiaozhu Liu and Shengxian Peng

914 Analysis of the Impacts on the Psychological Changes of Chinese Returning College Students After the Outbreak of the 2019 Coronavirus Disease
Yingying Xue, Kwak Han Pyong, Sae Sook Oh, Yingying Tao and Taofeng Liu

923 Application Research of Public Art in the Design of Old People's Rehabilitation Space Based on Virtual Information Interaction Platform From the Perspective of Developing Economies
Shaoqing Zhang

932 A Prognostic Model for Colon Adenocarcinoma Patients Based on Ten Amino Acid Metabolism Related Genes
Yangzi Ren, Shangwen He, Siyang Feng and Wei Yang

945 A Prediction Model for Peak Expiratory Flow Derived From Venous Blood Biomarkers and Clinical Factors in Amyotrophic Lateral Sclerosis
Xianghua He, Jiaming Feng, Xue Cong, Hongyan Huang, Quanzhen Zhao, Qiuyan Shen, Fang Xu and Yanming Xu

954 A Retrospective Study of Climate Change Affecting Dengue: Evidences, Challenges and Future Directions
Surbhi Bhatia, Dhruvisha Bansal, Seema Patil, Sharnil Pandya, Qazi Mudassar Ilyas and Sajida Imran

970 Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired
Surbhi Bhatia, Ajantha Devi, Razan Ibrahim Alsuwailem and Arwa Mashat

980 Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography
Fatemeh Sharifonnasabi, Noor Zaman Jhanjhi, Jacob John, Peyman Obeidy, Shahab S. Band, Hamid Alinejad-Rokny and Mohammed Baz

994 Analysis of the Impact of Supply Chain Relationship Strategy on Corporate Cash Dividend Policy Based on Chinese Data
Xingquan Yang and Lingyan Sun

1007 A Study on the Effect of the Pre-Go-Live Training in Anxiety and Depression of Medical Staff Based on the Data of Wuhan Fangcang Shelter Hospital During COVID-19 in the Era of Big Data
Qiang Feng, Huizhi Zhou, Lu Wang and Chuanyuan Kang

1013 The Effect of Big Data-Based Digital Payments on Household Healthcare Expenditure
Chengming Li, Daming Li, Si He, Shiqi Sun, Yuan Tian and Zeyu Wang

1025 Research on Infant Health Diagnosis and Intelligence Development Based on Machine Learning and Health Information Statistics
Siyu Wang, Min Li and Soo Boon Ng

1036 Analysis on the Cognitive Impact of Social Mobile Games on Left-Behind Children in the Era of Big Data
Mengyu Li, Jinglei Li, Megat Al Imran Yasin, Mohd Nizam Osman, Norliana Binti Hashim, Lay Hoon Ang and Yitian Xue

1046 Digital Analysis of Smart Registration Methods for Magnetic Resonance Images in Public Healthcare
Tao Chen, Mengxue Yuan, Jiajie Tang and Long Lu

1057 A Neural Network and Optimization Based Lung Cancer Detection System in CT Images
Chapala Venkatesh, Kadiyala Ramana, Siva Yamini Lakkisetty, Shahab S. Band, Shweta Agarwal and Amir Mosavi

1066 Effects of Osimertinib Combined With Pulmonary Rehabilitation and Health Care Training on Pulmonary Function, Complications, and Quality of Life in Patients After Radical Resection of Lung Cancer
Haijiang Xu, Ruixia Guo and Yantao Yang

1074 Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images
Ahila A, Poongodi M, Sami Bourouis, Shahab S. Band, Amir Mosavi, Shweta Agrawal and Mounir Hamdi

1088 SperoPredictor: An Integrated Machine Learning and Molecular Docking-Based Drug Repurposing Framework With Use Case of COVID-19
Faheem Ahmed, Jae Wook Lee, Anupama Samantasinghar, Young Su Kim, Kyung Hwan Kim, In Suk Kang, Fida Hussain Memon, Jong Hwan Lim and Kyung Hyun Choi

1105 Perception Exploration on Robustness Syndromes With Pre-processing Entities Using Machine Learning Algorithm
Pravin R. Kshirsagar, Hariprasath Manoharan, Shitharth Selvarajan, Hassan A. Alterazi, Dilbag Singh and Heung-No Lee

1119 Suggestions on Relieving Physical Anxiety of Medical Workers and Improving Physical and Mental Health Under the COVID-19 Epidemic—A Case Study of Meizhou City
Zhi-Cheng Lu, Chien-Hsing Tseng, Hsiao-Hsien Lin, Yuan-Shing Perng and Yi-Han Tseng

1134 Utility of Scoring System for Screening and Early Warning of Cervical Cancer Based on Big Data Analysis
Dan Hou, Binjie Yang, Yangdan Li and Ming Sun

1141 VCNet: Hybrid Deep Learning Model for Detection and Classification of Lung Carcinoma Using Chest Radiographs
Ritu Tandon, Shweta Agrawal, Arthur Chang and Shahab S. Band

1154 XGBoost-Based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers
Xianbin Song, Jiangang Zhu, Xiaoli Tan, Wenlong Yu, Qianqian Wang, Dongfeng Shen and Wenyu Chen

1160 A Survey on Machine Learning and Internet of Medical Things-Based Approaches for Handling COVID-19: Meta-Analysis
Shahab S. Band, Sina Ardabili, Atefeh Yarahmadi, Bahareh Pahlevanzadeh, Adiqa Kausar Kiani, Amin Beheshti, Hamid Alinejad-Rokny, Iman Dehzangi, Arthur Chang, Amir Mosavi and Massoud Moslehpour

1182 Mechanism of Negative Emotions of the Elderly in Normalization Period of COVID-19: A Mediated Mediation Model
Kai Xu

1190 Performance Evaluation of the Deep Learning Based Convolutional Neural Network Approach for the Recognition of Chest X-Ray Images
Sandhya Sharma, Sheifali Gupta, Deepali Gupta, Junaid Rashid, Sapna Juneja, Jungeun Kim and Mahmoud M. Elarabawy

1199 A Study on the Correlation Between Age-Related Macular Degeneration and Alzheimer's Disease Based on the Application of Artificial Neural Network
Meng Zhang, Xuewu Gong, Wenhui Ma, Libo Wen, Yuejing Wang and Hongbo Yao

1211 Breast Cancer Detection and Classification Empowered With Transfer Learning
Sahar Arooj, Atta-ur-Rahman, Muhammad Zubair, Muhammad Farhan Khan, Khalid Alissa, Muhammad Adnan Khan and Amir Mosavi

1229 Using Machine Learning Methods to Predict Bone Metastases in Breast Infiltrating Ductal Carcinoma Patients
Wen-Cai Liu, Ming-Xuan Li, Shi-Nan Wu, Wei-Lai Tong, An-An Li, Bo-Lin Sun, Zhi-Li Liu and Jia-Ming Liu

1242 A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients Undergoing Surgery With Prostate Cancer: A Population-Based Study
Zhaoxia Zhang, Chenghao Zhanghuang, Jinkui Wang, Tao Mi, Jiayan Liu, Xiaomao Tian, Liming Jin and Dawei He

1253 Intelligent automatic sleep staging model based on CNN and LSTM
Lan Zhuang, Minhui Dai, Yi Zhou and Lingyu Sun

1269 Interpretable clinical visualization model for prediction of prognosis in osteosarcoma: a large cohort data study
Wenle Li, Genyang Jin, Huitao Wu, Rilige Wu, Chan Xu, Bing Wang, Qiang Liu, Zhaohui Hu, Haosheng Wang, Shengtao Dong, Zhi-Ri Tang, Haiwen Peng, Wei Zhao and Chengliang Yin

1284 Big data-driven intelligent governance of college students' physical health: System and strategy
Chenliang Deng, Qiaoyan Yu, Ganglin Luo, Zhangzhi Zhao and Yuchao Li

1296 An efficient modular framework for automatic LIONC classification of MedIMG using unified medical language
Surbhi Bhatia, Mohammed Alojail, Sudhakar Sengan and Pankaj Dadheech

1317 Influence of seamless nursing model of humanistic care on nursing quality and blood infection risk of neurosurgical ICU patients and its correlation with APACHE II score
Fuhua Liu and Xiaoting Wang

1325 Retinal Vessel Segmentation Based on B-COSFIRE Filters in Fundus Images
Wenjing Li, Yalong Xiao, Hangyu Hu, Chengzhang Zhu, Han Wang, Zixi Liu and Arun Kumar Sangaiah

1339 A home fitness satisfaction model for Chinese residents during the COVID-19 pandemic based on SEM analysis
Dexu Chen and Bin Zhang

frontiers in Public Health

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/fpubh.2021.729795

Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate

Jayakumar Kaliappan[1], *Kathiravan Srinivasan*[1], *Saeed Mian Qaisar*[2], *Karpagam Sundararajan*[3], *Chuan-Yu Chang*[4]* *and Suganthan C*[5]

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, [2] Electrical and Computer Engineering Department, Effat University, Jeddah, Saudi Arabia, [3] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [4] Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu, Taiwan, [5] School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Mohamed Hammad,
University of Menoufia, Egypt
Joseph Henry Arinze Anajemba,
Hohai University, China
Muhammad Fazal Ijaz,
Sejong University, South Korea
Ebuka Ibeke,
Robert Gordon University,
United Kingdom

***Correspondence:**
Chuan-Yu Chang
chuanyu@yuntech.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 23 June 2021
Accepted: 16 August 2021
Published: 14 September 2021

Citation:
Kaliappan J, Srinivasan K, Mian Qaisar S, Sundararajan K, Chang C-Y and C S (2021) Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate. Front. Public Health 9:729795. doi: 10.3389/fpubh.2021.729795

This paper aims to evaluate the performance of multiple non-linear regression techniques, such as support-vector regression (SVR), k-nearest neighbor (KNN), Random Forest Regressor, Gradient Boosting, and XGBOOST for COVID-19 reproduction rate prediction and to study the impact of feature selection algorithms and hyperparameter tuning on prediction. Sixteen features (for example, Total_cases_per_million and Total_deaths_per_million) related to significant factors, such as testing, death, positivity rate, active cases, stringency index, and population density are considered for the COVID-19 reproduction rate prediction. These 16 features are ranked using Random Forest, Gradient Boosting, and XGBOOST feature selection algorithms. Seven features are selected from the 16 features according to the ranks assigned by most of the above mentioned feature-selection algorithms. Predictions by historical statistical models are based solely on the predicted feature and the assumption that future instances resemble past occurrences. However, techniques, such as Random Forest, XGBOOST, Gradient Boosting, KNN, and SVR considered the influence of other significant features for predicting the result. The performance of reproduction rate prediction is measured by mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R-Squared, relative absolute error (RAE), and root relative squared error (RRSE) metrics. The performances of algorithms with and without feature selection are similar, but a remarkable difference is seen with hyperparameter tuning. The results suggest that the reproduction rate is highly dependent on many features, and the prediction should not be based solely upon past values. In the case without hyperparameter tuning, the minimum value of RAE is 0.117315935 with feature selection and 0.0968989 without feature selection, respectively. The KNN attains a low MAE value of 0.0008 and performs well without feature selection and with hyperparameter tuning. The results show that predictions performed using all features and hyperparameter tuning is more accurate than predictions performed using selected features.

Keywords: COVID-19, feature selection, machine learning, prediction error, reproduction rate prediction, regression

INTRODUCTION

The world has witnessed several deadly diseases at different times. In the year 2020, the world suffered a serious pandemic that took away many lives (1). The coronavirus disease (COVID-19) is a disease that started as an epidemic and evolved into a pandemic. The disease was first discovered in late December 2019 in Wuhan, China (2). The virus responsible for causing the disease is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is highly contagious and causes severe respiratory issues. The virus rapidly spread across the world and affected 223 countries, infected more than 9.3×10^7 people, and took over 2×10^6 human lives, according to the WHO report in January 2021 (3). As a result, scientists and epidemiologists worldwide are investigating the virus to reduce its impact on human lives.

The coronavirus is named after the word, "coronation" since the spikes on the surface of the virus resemble a crown. This virus was believed to be an animal virus in 2002. The SARS-CoV is mostly found in bats and transmitted to other animals, such as cats. The first human-infected coronavirus case was reported in 2003 in Guangdong province in the south of China (4).

Knowledge of the immune system of our body is required to understand the mechanisms of the COVID-19 or any other viral infections. Viruses are microorganisms that make our body cells their hosts for replication and multiplication. The immune system of our body is activated by the entry of the virus and identifies the virus as an alien body for destruction. After attacking and killing the viruses, the immune system "remembers" the virus and launches the same protective measures when the virus enters again. Viruses are capable of fast evolutions. They evolve to new shapes or mechanisms to survive in the changing environment.

Viral infections often affect people with weak immune systems. The elderly, children, and people with medical conditions are prone to the attack of novel viruses. The virus can be deadly and threatening to the senior population, especially the elderly with chronic medical conditions.

The SARS-CoV-2 is transmitted *via* respiratory droplets expelled by sneezing, coughing, or talking. The virus can also be contracted by touching a contaminated surface. One significant property of the SARS-CoV-2 is its capacity to survive on various surfaces for up to 9 days at room temperature, which facilitates its rapid transmission (5). Acute Respiratory Disease Syndrome is caused predominantly by this virus and often leads to multiple organ dysfunctions, resulting in physiological deterioration and even death of the infected persons (6).

This study is intended to predict the rate of reproduction of the deadly SARS-CoV-2. The reproduction rate (R_o) is an important parameter to predict the spread of a disease in a pandemic situation. The R_o value indicates the transmissibility of a virus through the average number of new infections caused by an infectious person in a naïve population. The value of $R_o < 1$ indicates that the infection would die out. On the other hand, if the value is >1, the spread of the disease would increase. For example, a reproduction rate of 18 indicates that a single infected individual can potentially infect 18 more individuals. The reproduction rate is needed to determine whether the disease is under control or turning into an epidemic.

There are many standard methods to predict the reproduction rate. The XGBoost is an optimal Gradient Boosting algorithm with tree pruning, parallel processing, missing value handling, and by the efficient use of hardware resources and regularization to avoid overfitting and bias. The XGBoost has faster computational times (7) in all types of environments. The XGBoost is an improvement on the Gradient Boosting algorithm. Training models with an XGBoost iterative boosting approach remove errors at preceding boosting trees in the following iterations (8). Support vector regression (SVR) was based on the Vapnik–Chervonenkis (VC) theory. It is used when the output is a continuous numerical variable. Support vectors are data points closest to the hyperplanes. The hyperplanes represent the decision boundaries. The Radial Basis Function is a commonly used kernel function. The use of a kernel function is to transform the data into a higher-dimensional space. The SVR and convolutional neural network (CNN) have been used to detect groundwater location, and comparative results showed that the SVR outperforms the CNN (9). In k-nearest neighbor (KNN), the outcome of the variable is determined by taking an average of the observations found in the same neighborhood. The KNN algorithm assigns a weight, "W" to the KNN and a weight of 0 to the others.

The performance of the machine learning algorithms depends on the hyperparameter values. The values for the hyperparameters can be assigned in three ways:

1) Based on default values given in the software packages.
2) Manually configured by the user.
3) Assigned by algorithms, such as the simple grid search, random search, Bayesian optimization, ANOVA approach, and bio-inspired optimization algorithms.

The process of identifying the most relevant features is referred to as "feature selection." The three main advantages of feature selection are:

- Simplifying the interpretation of the model.
- Reducing the variance of the model to avoid overfitting.
- Reducing the computational cost (and time) for model training.

Motivation

Artificial intelligence (AI) has been successful in many fields and facilitates our daily life in various ways (10–17). The reproduction rate prediction is crucial in successfully establishing public healthcare in the battle against COVID-19. The prediction of the reproduction rate is performed by using not just the past values but also by using the closely related factors. This work also investigates the impact of feature selection and hyperparameter tuning on the performance of non-linear machine learning techniques.

Research Gap and Contribution

The reproduction rate is related to many factors, such as the average number of contacts a person has, number of days a

person is infected, from the day of exposure to the disease, the number of active cases, values of stringency index, testing capacity and positivity, and so on. As a result, the reproduction rate, time curve, and future values cannot be satisfactorily estimated by the probability distribution functions alone (18).

Time series prediction models, such as the autoregressive integrated moving average (ARIMA), Gray Model, and Markov Chain models do not consider multiple factors in reproduction rate prediction. Autoregressive models assume that future values resemble the past. Mechanistic models based on the susceptible, exposed, infected, and recovered (SEIR) states framework or modified version of the framework use the time series data to hold the currently confirmed cases, removed cases (including recovered and deceased cases), and time-varying transmission rates. However, some factors are still not included, and there is no weighting for the factors (19).

So there arises a need to study the various factors acting on the reproduction rate and to prioritize it. Hence identifying the importance of various features (for example, Total_cases_per_million, Total_deaths_per_million) under factors like active cases, stringency index, testing capacity, and positivity are done using feature selection algorithms. Multiple regression uses several explanatory variables to predict the single response variable. The performance of the non-linear machine learning techniques, such as Random Forest, XGBOOST, Gradient Boosting, KNN, and SVR are used in reproduction rate prediction. The performance of these approaches for predicting the COVID-19 reproduction rate was measured using the evaluation metrics like mean absolute error (MAE), mean squared error (MSE), **r**oot mean squared **e**rror (RMSE), R-Squared, relative absolute **e**rror (RAE), and root relative squared error (RRSE). The influence of feature selection and hyperparameter tuning operation on their performance is also studied.

Structure of the Paper

Section Introduction of the paper introduces the reproduction rate, feature selection, machine learning techniques and hyperparameters, and the motivation of this study. Section Related Works discusses the related works, as well as the identified research gap and the contributions. Section Materials and Methods describes the methods used in this work, including feature selection, hyperparameter tuning, and prediction and performance measurement. Section Results and Discussion discusses the experimental results. Finally, section Conclusion and Future Work provides the conclusion and future work.

RELATED WORKS

Zivkovic et al. (20) proposed a hybridized machine learning, adaptive neuro-fuzzy inference system with enhanced beetle antennae search (BAS) swarm intelligence metaheuristics. The results showed that the system is a good prediction model in time series forecasting. The defects in the BAS algorithm were rectified using the Cauchy exploration strategy BAS (CESBAS) using the Cauchy mutation and three additional control parameters. The selection of optimum values for the adaptive network-based fuzzy inference system (ANFIS) parameters became an NP-hard optimization problem. The ANFIS parameters values were determined using the CESBAS metaheuristics algorithm. The performance metrics, such as RMSE, MAE, MAPE, RMRE, and R-Squared, were used to evaluate the outcomes on influenza datasets.

The research goal in Mojjada et al. (21) was to forecast the number of new COVID-19 cases, mortalities, and recoveries using various machine learning regression models, such as the lowest absolute and selective shrinking operator (LASSO), vector supports, such as short message service (SMS), and exponential smoking (ES) models. While the linear regression and LASSO models were more effective in estimating and verifying the death rate, the ES model provided the overall best results.

Farooq and Bazaz (22) used an artificial neural network (ANN) to forecast the COVID-19 based on an online incremental learning technique using an adaptive and non-intrusive analytical model. The COVID-19 data was updated every day, so online incremental learning was the best option for forecasting since there is no need to retrain or rebuild the model from scratch.

Milind et al. (23) discovered many factors behind the spread of the coronavirus, such as the relationship between the weather and the spread of COVID-19, growth rate, and mitigation. Support vector regression (SVR) was used to predict the transmission rate, epidemic end, and the spread of the coronavirus across regions, and to analyze the growth rates and the types of mitigation across countries. The Pearson coefficient was used in the correlation between the coronavirus and weather correlation coefficient. Weather factors, such as the wind speed, temperature, and humidity of Milan city in Italy and New York City in the United States were considered. The SVR is a non-parametric technique since it only depends on the kernel function, implying that there is no need to change the explanatory variables in constructing a non-linear model. The study also compared the performances of SVR, linear regression, and polynomial regression.

Chicco and Jurman (24) predicted the survival of patients who had heart failure based on the ejection fraction and serum creatinine level. A database of 299 patients collected in 2015 was used. Feature selection was performed, and the factors were ranked. The ejection fraction and serum creatinine levels were found to be highly relevant among the 13 selected features. As a result, the prediction model was built and executed based on these two factors.

Mortazavi et al. (25) investigated the capability of machine learning techniques when applied to a high dimensional and non-linear relationship. They predicted the readmission of patients hospitalized for heart failure. The prediction was performed with various machine learning techniques, such as Random Forest, Gradient Boosting, and Random Forest combined hierarchically with support vector machines (SVMs) or logistic regression (LR) and Poisson regression. The obtained results were tested against traditional LR methods. The model was evaluated using the receiver operating characteristics (ROC) curve (C statistic), the positive predictive value (PPV), sensitivity, specificity, and

f-score. The ROC was found to be a good measure for model discrimination.

Balli (26) analyzed the COVID-19 data from Germany, the United States, and other parts of the world. Methods including the SVM, linear regression, multilayer perceptron, and Random Forest methods were used to model the COVID-19 data. The performances of the methods were compared using the RMSE, absolute percentage error (APE), and mean absolute percentage error (MAPE). Among the tested methods, the SVM outperformed all other methods in the COVID-19 data modeling and was successfully used to diagnose the behavior of cumulative COVID-19 data over time.

A system to handle the data with non-linear relationships and non-normal distribution was proposed by Kuo and Fu (27). A total of 52 input variables relating to confirmed cases, environment variables, country-dependent variables, community mobility variables, and time series variables were used in the study (27). The impact of population mobility had caused an increase in the number of infections over the weekend. This work served as a basis for researchers analyzing geographical characteristics, seasonality, as well as models, such as long short-term memory (LSTM), ARIMA, convolutional neural network (CNN), and so on.

The COVID Patient Detection System (CPDS) used by Shaban et al. (28) was designed using a Hybrid Feature Selection Method (HFSM) consisting of two stages, a fast selection stage (FS^2) and an accurate selection stage (AS^2). The FS^2 used several filter methods, and the filtered features served as the initial population of the genetic algorithm, which was used as a wrapper method. An enhanced K-nearest neighbor (EKNN) classifier was used to solve the trapping problem. The most significant features from the chest CT images of patients were selected. The HFSM allowed the EKNN classifier to obtain rapid predictions with high accuracy. The proposed feature selection algorithm was compared with four recent feature selection techniques, and the proposed CPDS had achieved an accuracy of 96%.

Sujatha et al. (29) utilized the linear regression, multi-layer perceptron (MLP), and vector autoregression (VAR) models to foresee the spread of the COVID-19 using the COVID-19 Kaggle data. The correlations between the features of the dataset are crucial in finding the dependencies. The VAR model is a more suitable analysis model for multivariate time series. It is an m-equation, m- variable model where an individual variable is based on its current and past values. The MLP methods provide better predictions than the linear regression and VAR models.

Yang et al. (30) predicted the number of new confirmed cases using SEIR and AI methods. The authors used the probability of transmission, incubation rate, and the probability of recovery or death as factors in the predictions. New case predictions made by the AI method are more accurate than the SEIR predictions.

The Gradient Boosting Feature Selection (GBFS) algorithm learns the ensemble of regression trees to identify the non-linear relationship between features with ease. The classification error rates for the GBFS and Random Forest methods are the lowest, whereas the L1-regularized logistic regression (L1-LR) and Hilbert–Schmidt independence criterion (HSIC) Lasso methods have higher error rates (31).

The XGBOOST algorithm was applied to calculate the business risk by Wang (32). Several feature selection methods were used to find the redundant features. Two hyper-parameter

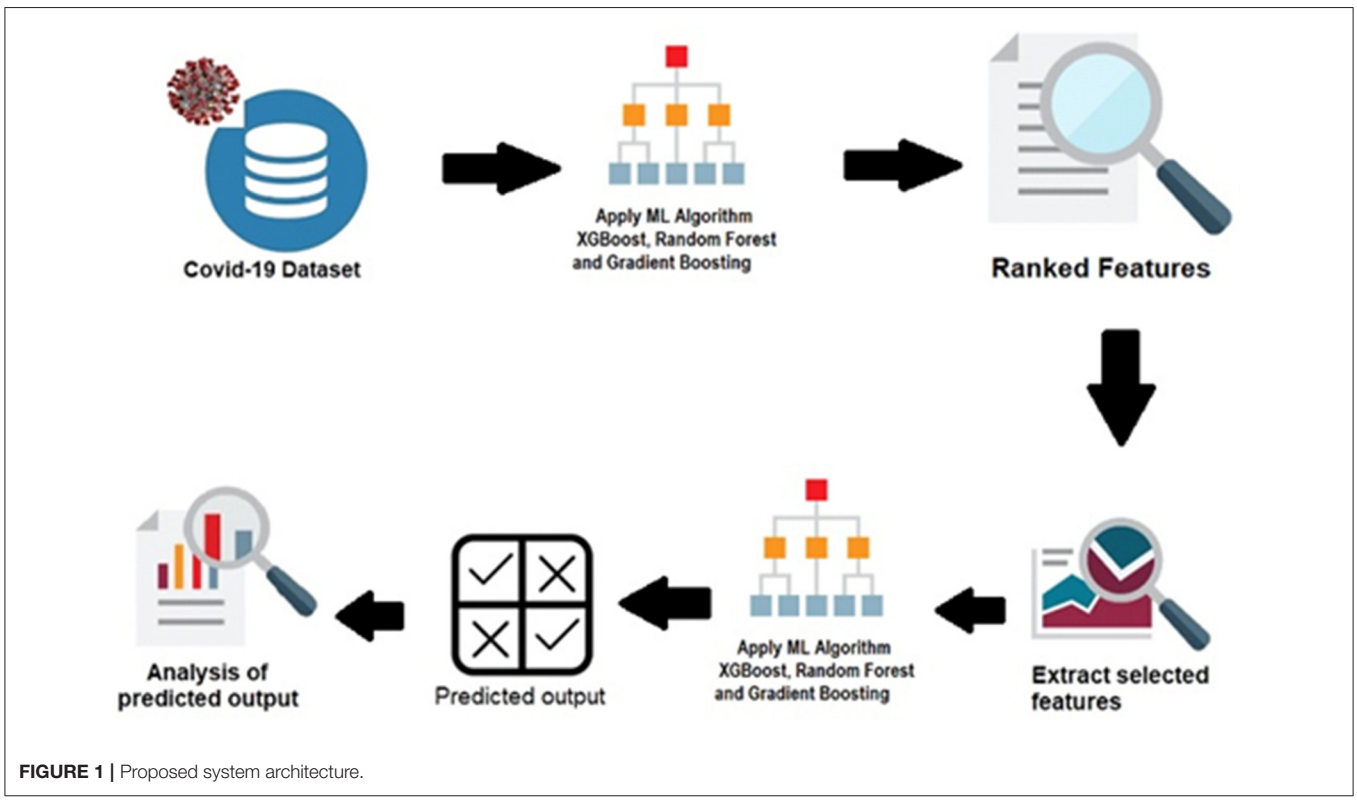


FIGURE 1 | Proposed system architecture.

optimization approaches were applied: random search (RS) and Bayesian tree-structured Parzen Estimator (TPE). The XGBOOST with hyper-parameter optimization performed well for business risk modeling.

Chintalapudi et al. (33) used the predicted reproduction rate to forecast the daily and the cumulative COVID-19 cases for the next 30 days in Marche, Italy. The probability-based prediction was performed with the maximum likelihood function. In the implementation, a simple linear regression method was used to fit the exponential growth of infected incidences over time, and the linear regression was applied over the incidence data. This study showed that the outbreak size and daily incidence are primarily dependent on the daily reproductive number.

Locatelli et al. (34) estimated the COVID-19 reproduction rate of Western Europe with the average from 15 countries. The authors used the generation interval, defined as the time needed for an infected person to infect another person and for reproduction rate estimation. The works by Zhang et al. (35) and by Srinivasu et al. (36), Panigrahi et al. (37, 38), Tamang (39), Chowdhary et al. (40), and Gaur et al. (41) demonstrated the efficacy of machine learning algorithms in various fields.

TABLE 1 | Feature Importance Scores obtained using Random Forest Regression, XGBoost, and Gradient Boosting.

Features	Feature	Random forest regression feature importance	XGBOOST feature importance	Gradient boosting feature importance
Feature: 0	Total_cases	0.10044	0.92185	0.03024
Feature: 1	New_cases	0.02539	0.00055	0.00058
Feature: 2	Total_deaths	0.08816	0.00011	0.10000
Feature: 3	New_deaths	0.01286	0.00209	0.00000
Feature: 4	Total_cases_per_million	0.10196	0.00013	0.07758
Feature: 5	New_cases_per_million	0.01562	0.00023	0.00176
Feature: 6	Total_deaths_per_million	0.09868	0.00011	0.11893
Feature: 7	New_deaths_per_million	0.00709	0.00021	0.00003
Feature: 8	New_tests	0.00253	0.04245	0.00006
Feature: 9	Total_tests	0.04384	0.00012	0.08950
Feature: 10	Total_tests_per_thousand	0.06210	0.00000	0.05340
Feature: 11	New_tests_per_thousand	0.01958	0.00000	0.00022
Feature: 12	Positive_rate	0.00394	0.00030	0.00090
Feature: 13	Tests_per_case	0.00307	0.00019	0.00376
Feature: 14	Stringency_index	0.01744	0.00020	0.05846
Feature: 15	Population_density	0.00000	0.00000	0.00000

MATERIALS AND METHODS

The spread of the COVID-19 depends on many factors. New factors influencing the spread of the disease are still being discovered, and the identification of predominant factors is crucial. The prediction of COVID-19 spread is highly related to the feature-reproduction rate. Data science can be applied to track the crucial features used for the prediction from any number of features. Traditional statistical approaches, such as the chi-square and Pearson correlation coefficient provide the importance of the features in relation to the other features. Feature selection reduces the overfitting and underfitting problems, computational cost, and time. The reproduction rate prediction is important since it is associated with the status of the COVID-19. Feature-ranking is performed using Random Forest regression, Gradient Boosting, and XGBoost. Seven factors are considered in this study: the total number of cases, number of new cases, total number of deaths, total number of cases per million, total number of deaths per million, total number of tests conducted per thousand, and the positive rate. The proposed system architecture is represented in **Figure 1**.

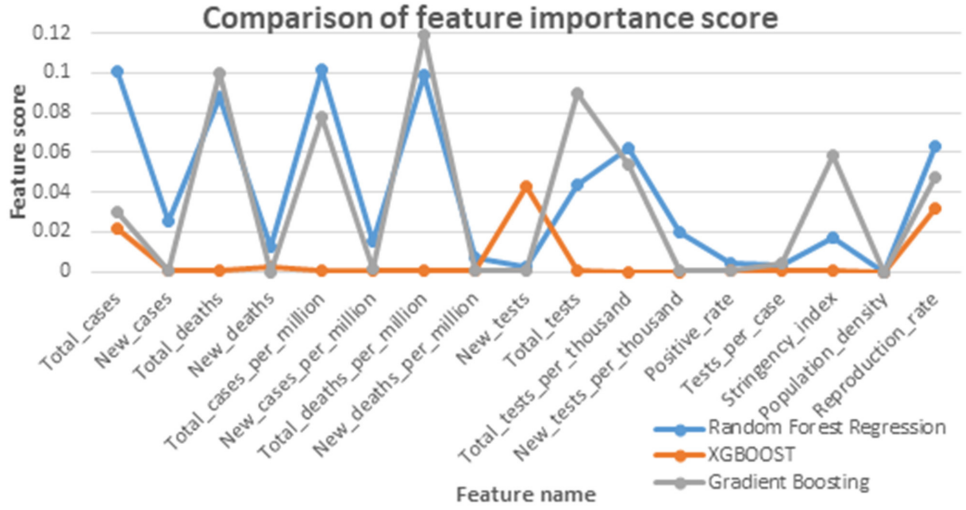


FIGURE 2 | Comparison graph of feature score given by feature selection algorithms.

Reproduction Rate

Newly occurring diseases can be detrimental for humans and other animals, whether the diseases are caused by a new pathogen or a modified form of an existing pathogen (19). In this work, the simple compartmental disease models and matrix methods are used to calculate the reproduction rate, R0.

Feature Selection

Embedded filter-based feature selection methods, such as Random Forest, Gradient Boosting, and XGBoost, which take into account the regression process, are used in this work. The Random Forest approach is an embedded feature selection method in which hundreds of decision trees are constructed by extracting random observation values of random features. The training determines features that reduce impurity. The principle of Gradient Boosting, and XGBoost methods are used to boost the weak learners. Gradient Boosting strives to minimize the error between the predicted and the actual values. The XGBoost is an extreme Gradient Boosting algorithm. The XGBoost is the regularized form of Gradient Boosting (42). The XGBoost is fast with L1, L2 regularization and parallel computing. It delivers high performance since it works on the second partial derivatives of the loss function. The main highlight of the Random Forest algorithm lies in its ability to prevent overfitting and increase accuracy. The advantage of gradient boosting is its ability to tune many hyperparameters and loss functions.

Parameter Settings

The experiments use multiple non-linear regression tree algorithms and the result is implemented in Python. For experiments without hyperparameter tuning, the default values in the SciKit library are used. For Random Forest regression, the parameter values are initialized as follows: n_estimators = 100, n_jobs = −1, oob_score = True, bootstrap = True, and random_state = 42. For XGBoost, the XGB Regressor method is used to fit the test data, and n_estimatorsis set to 100. The Gradient Boosting feature importance is calculated by setting the value of n_estimators to 500, max_depth to 4, min_samples_split to 5, learning_rate to 0.01, and loss as ls. For the KNN algorithm, the lower error rate is achieved when the K value equals 7. For the SVR, the radial basis function kernel is used with degree = 3 and gamma = scale. For experiments with hyperparameter tuning, grid search and randomized approaches are used. A grid search exhaustively tests all possible combinations of the specified hyperparameter values for an estimator. In a randomized search, the model selects the combinations randomly. Both approaches are very effective ways of tuning the parameters to increase the generalizability of the model. The GridSearchCV method of sklearn tunes the hyperparameters of the SVR, KNN, XGBoost, and Gradient Boosting approaches. The randomized search CV function is used for the hyperparameter tuning of Random Forest Regressor.

Dataset

The dataset was taken from the website "https://github.com/owid/covid-19-data/tree/master/public/data" (43). A total of 16 fields were used for the study of reproduction rate. They are Total_cases, New_cases, Total_deaths, New_deaths, Total_cases_per_million, New_cases_per_million, Total_deaths_per_million, New_deaths_per_million, New_tests, Total_tests, Total_tests_per_thousand, New_tests_per_thousand, Positive_rate, Tests_per_case, Stringency_index, Population_density. Records from April 1, 2020, to November 30, 2020, are used as training data (244 records/day). Records from December 1, 2020, to March 10, 2021, are used as testing data (100 records/day).

Performance Metrics

Numerous machine learning (ML)-based predictive modeling techniques are used in the COVID-19 predictions. Therefore, there is a need to measure the performance of each model and its prediction accuracy. The metrics used to assess the effectiveness

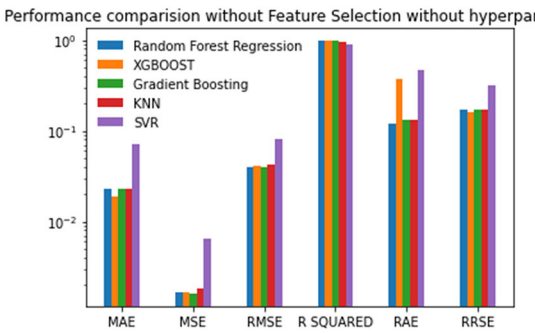


FIGURE 3 | Performance comparison graph of regression techniques without feature selection and without hyperparameter tuning.

TABLE 2 | Prediction without Feature Selection and without hyperparameter tuning.

Sl. No	Performance metrics	Prediction without feature selection and without hyperparameter tuning				
		Random forest regression	XGBOOST	Gradient boosting	KNN	SVR
1	MAE	0.0230122	**0.0189412**	0.02226608	0.0228918	0.0712651
2	MSE	0.0016347	0.0016482	**0.00135535**	0.0018072	0.0064267
3	RMSE	0.0404316	0.0405992	**0.03681510**	0.0425122	0.0801667
4	R-Squared	**0.9792338**	0.9790759	0.97830657	0.9710729	0.8971356
5	RAE	0.1206129	0.3754830	**0.11731593**	0.1306129	0.4754830
6	RRSE	0.1700794	0.1605438	**0.14728681**	0.1700794	0.3207246

TABLE 3 | Prediction with Feature Selection and without hyperparameter tuning.

Performance metrics	Prediction with feature selection and without hyperparameter tuning				
	Random forest regression	XGBOOST	Gradient boosting	KNN	SVR
MAE	0.02168571	**0.0183910**	0.0225887	0.024706	0.0902574
MSE	**0.00139726**	0.0016258	0.0016091	0.001867	0.0090748
RMSE	**0.03738006**	0.0403218	0.0401140	0.043210	0.0952621
R-Squared	0.97415592	0.9796126	**0.9798852**	0.970115	0.8547498
RAE	0.11583870	**0.0968989**	0.1301729	0.120172	0.3755501
RRSE	**0.15999851**	0.1613162	0.1728719	0.162871	0.1728719

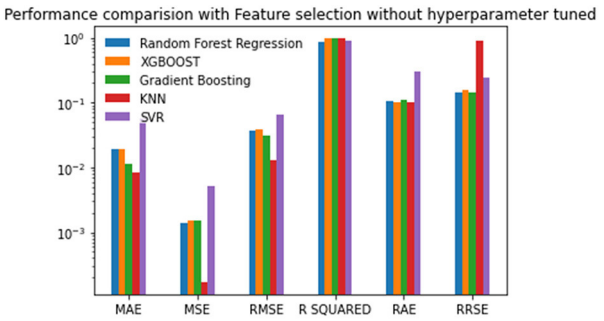


FIGURE 4 | Performance comparison graph of regression techniques with feature selection algorithm and without hyperparameter tuning.

of the model in predicting the outcome are very important since they influence the conclusion. The performance metrics to identify the error rate between the predicted and observed values are as follows:

- Root mean square error (RMSE)
- Mean absolute error (MAE)
- Determination coefficient (R^2)
- Relative absolute error (RAE)
- Root relative squared error (RRSE)

Mean Absolute Error

The mean absolute error measures the sum of the absolute differences between the predicted output and the actual output. One cannot identify whether it is under predicting or over predicting since all variations have equal weight.

Equation 1 provides the formula to calculate the MAE.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|SWL_{FOR,i} - SWL_{OBS,i}|, \tag{1}$$

where $SWL_{FOR,i}$ represents the forecast output, $SWL_{OBS,i}$ represents the actual output, N represents the total number of data points, and I represents a single data entry from the data points.

Root Mean Squared Error

The RMSE measures the square root of the average squared deviation between the forecast and the actual output, as given in Equation 2. It is used when the error is highly non-linear. The RMSE indicates the amount of errors in the predicted data on average and is a good measure of the prediction accuracy.

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(SWL_{FOR,i} - SWL_{OBS,i})^2}{N}} \tag{2}$$

Determination Coefficient

The R^2 metric shows the percentage variation in y explained by x-variables, where x and y signify a set of data. It finds the likelihood of the occurrence of a future event' in the predicted outcome, as given in Equation 3.

$$R^2 = \left(\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}\right)^2 \tag{3}$$

Relative Absolute Error

Relative Absolute Error (RAE) metric gives the ratio of residual or mean error to the forecast error of a naive model. Equation 4, returns a value less than 1, when the proposed model performs better than the naïve model. In Equation 4, "P" stands for the predicted value and "A" for the actual value.

$$RAE = \frac{\left[\sum_{i=1}^{n}(P_i - A_i)^2\right]^{\frac{1}{2}}}{\left[\sum_{i=1}^{n}A_i^2\right]^{\frac{1}{2}}} \tag{4}$$

Root Relative Squared Error

The Root Relative Squared Error (RRSE) is given as the square root of the relative squared error (RSE). The RSE metric compares the actual forecast error to the forecast error of a naive model. It can be used in models whose errors are measured in different units. As given in Equation 5 and 6, the total squared error is divided by the total squared error of the simple predictor. The simple predictor is just the average of the actual values. The predicted output is "P" and "T" is the target value. Further, the value "i" represents the model and j represents the record.

RRSE is given as the square root of the relative squared error, which provides the error in the dimensions of the quantity being predicted, as given in Equation (7). RSE_i represents the relative squared error for the model "i".

$$RSE_i = \frac{\sum_{j=1}^{n} (P_{ij} - T_j)^2}{\sum_{j=1}^{n} (T_j - \overline{T})^2} \quad (5)$$

$$\overline{T} = \frac{1}{n} \sum_{j=1}^{n} T_j \quad (6)$$

$$RRSE = \sqrt{RSE_i} \quad (7)$$

RESULTS AND DISCUSSION

Results

All experiments were performed using Python's Sci-Kit Library on a Jupyter notebook. The feature importance scores obtained by Random Forest regression, XGBoost, and Gradient Boosting are given in **Table 1** and plotted in **Figure 2**.

Out of the 16 features, the top seven features affecting the reproduction rate are identified from the obtained feature importance scores. The seven features are Total_cases, New_cases, Total_deaths, Total_cases_per_million, Total_deaths_per_million, Total_tests, Total_tests_per_thousand, and Positive_rate.

Experiments were conducted for the reproduction rate prediction using Random Forest, XGBoost, Gradient Boosting, support-vector regression (SVR), and k-nearest neighbor (KNN) regression methods. In addition, the experiments were intended to investigate the impacts of feature selection and hyperparameter tuning. Four experiments were conducted with and without feature selection or hyperparameter tuning. Experiment 1 was conducted using the five regression techniques without feature selection and without parameter tuning. Experiment 2 was conducted with feature selection and without parameter tuning. Experiment 3 was conducted without feature selection and with hyperparameter tuning, and finally, Experiment 4 was conducted with feature selection and with parameter tuning. The reproduction rate prediction was measured using the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R-Squared, relative absolute error (RAE), and root relative squared error (RRSE). The best, the second-best and the worst results for the particular metrics and experiments are discussed in detail below. The best results obtained for the metrics are given in bold in the **Tables 2**, **3**, **5**, **6**.

The first experiment used all features in the reproduction rate prediction, and each of the regression techniques used the default values for the hyperparameter. The resulting performance metric values are given in **Table 2**. The Gradient Boosting method performs well with the lowest MSE, RMSE, RAE, and RRSE values and the second-best score for MAE. Random Forest is the next best algorithm with the best R-Squared value and the second-best scores for MSE, RMSE, and RAE. The XGBoost has an average performance. The SVR has the worst scores in all of the performance metrics. The lowest MAE of 0.0189412 was obtained by XGBoost followed by Gradient Boosting with the second-best MAE of 0.02226608. The SVR has the highest MAE of 0.0712651. The minimum MSE, RMSE, RAE, and RRSE values of 0.00135535, 0.036815107, 0.1173159354, and 0.1472868, respectively, are achieved by Gradient Boosting. Random Forest

TABLE 4 | Best tuned values of the hyperparameters for the different regression techniques.

ML algorithms	Hyper parameter values
Random Forest Regression	{"n_estimators": 800, "min_samples_split": 2, "min_samples_leaf": 1, "max_features": "auto," "max_depth": 100, "bootstrap": True}
KNeighbors Regressor	(n_neighbors = 3, weights = "distance")
Support Vector Regression	{"C": 1.5, "epsilon": 0.1, "gamma": 1e-07, "kernel": "linear"}
XGBOOST	{"colsample_bytree": 0.7, "learning_rate": 0.1, "max_depth": 5, "min_child_weight": 3, "n_estimators": 500, "objective": "reg:squarederror," "subsample": 0.5}
Gradient Boosting Regression	{"learning_rate": 0.02, "max_depth": 10, "n_estimators": 1,500, "subsample": 0.5}

TABLE 5 | Prediction without Feature Selection and with hyperparameter tuning.

Performance metrics	Prediction without feature selection and with hyperparameter tuning				
	Random forest regression	XGBOOST	Gradient boosting	KNN	SVR
MAE	0.021551531	0.020971233	0.021192539	**0.008626385**	0.064263084
MSE	0.001430578	0.001684012	0.001578488	**0.000225718**	0.005462202
RMSE	0.037822976	0.041036714	0.039730185	**0.015023919**	0.073906711
R-Squared	0.977102535	0.973046128	0.974735124	0.996387212	**0.912573368**
RAE	0.113551075	0.110493591	0.111659615	**0.045450843**	0.338590443
RRSE	0.151319084	0.164176344	0.158949287	**0.060106472**	0.295679949

achieves the maximum R-squared value of 0.97923. The obtained metric values are plotted in **Figure 3**.

The seven selected features were used in the second experiment, and the prediction process was performed without hyperparameter tuning. There is a reduction in the MAE, MSE, and RMSE values of about 0.01 using Random Forest regression. There is a marginal reduction in the MAE, MSE, and RMSE values using XGBoost and Gradient Boosting for feature selection. The lowest MAE value of 0.018391, RAE value of 0.096898, and the best R-squared value of 0.9796126 is achieved by XGBoost. Random Forest gives the lowest MSE of 0.00139, RMSE of 0.037380, and RRSE of 0.159998. Out of all the algorithms, the SVR produces the highest error rate. The results are given in **Table 3** and plotted in **Figure 4**. The Random Forest regression and XGBoost techniques performed better than the other techniques. The performance of SVR is the worst among the algorithms compared for reproduction rate prediction. In the experiment with feature selection and without hyperparameter tuning, the Random Forest approach has achieved the top performance with three best scores and two second-best scores. The XGBoost has the best scores for the MAE, R-squared, and RAE, and the second-best score for RRSE.

Experiment 3 was conducted without feature selection and with hyperparameter tuning. The tuned hyperparameter values are listed in **Table 4**. The results are good after hyperparameter tuning is performed with the grid search or random search. The results are analyzed based on the best, the second-best, and the last scores. The performance of KNN tops all of the other algorithms when the experiment is performed without feature selection and with hyperparameter tuning. Random Forest is the next best algorithm with the second-best scores for the MSE, RMSE, R-Squared and RRSE. The values are given in **Table 5** and plotted in **Figure 5**.

Experiment 4 was conducted with feature selection and with parameter tuning. In this experiment, the Random Forest approach has the best scores for the MSE, RMSE, R-Squared, and RRSE. The XGBOOST has the best scores for the MAE and RAE. Nevertheless, Gradient Boosting has the second-best scores in the MSE, RMSE, R-Squared, and RRSE. The KNN has two second-best scores. Again, the SVR has the worst scores for all of the performance metrics. The values are given in **Table 6** and plotted in **Figure 6**.

The computation times for the different types of prediction are computed and listed in **Table 7**. The Random Forest algorithm uses a random search technique for hyperparameter tuning, which requires more time. All of the other algorithms use the grid search technique. The KNN and SVR are able to perform hyperparameter tuning rapidly. XGBOOST and Gradient Boosting regression have moderate running times of around 100 s.

The predicted and actual reproduction rates for Random Forest, KNN, SVR, XGBoost, and Gradient Boosting are, respectively, plotted in **Figures 7–11**. The graphs show that the predicted values are very close to the actual values.

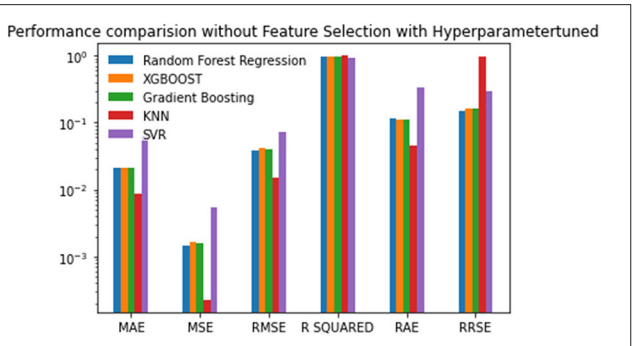


FIGURE 5 | Performance comparison graph of regression techniques without feature selection and with hyperparameter tuning.

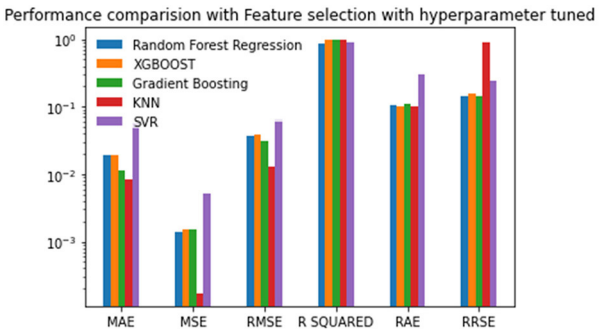


FIGURE 6 | Performance comparison graph of regression techniques with feature selection algorithm and with hyperparameter tuning.

TABLE 6 | Prediction with Feature Selection and with hyperparameter tuning.

Performance metrics	Prediction with feature selection and with hyperparameter tuning				
	Random forest regression	XGBOOST	Gradient boosting	KNN	SVR
MAE	0.020152332	**0.019385962**	0.021180387	0.019439098	0.076691979
MSE	**0.001476030**	0.001715576	0.001683149	0.001706222	0.007260537
RMSE	**0.038419141**	0.041419510	0.041026197	0.04130644	0.085208784
R-Squared	**0.976375028**	0.972540924	0.973059942	0.97269064	0.883789669
RAE	0.106178955	**0.102141089**	0.111595587	0.102421055	0.404076017
RRSE	**0.153704171**	0.165707803	0.164134267	0.165255439	0.340896364

TABLE 7 | Running time of algorithms with hyperparameter tuning and prediction.

ML algorithms	Running time including hyperparameter tuning with all features (in seconds)	Running time including hyperparameter tuning with selected feature (in seconds)
Random Forest Regression	492.46183967590330	383.7128930091858
KNeighbors Regressor	0.2534364700317383	0.131617689132690
Support Vector Regression	0.8226490020751953	0.512803316116333
XGBOOST	165.45190143585205	145.1453814506530
Gradient Boosting Regression	120.36712908744812	93.36355471611023

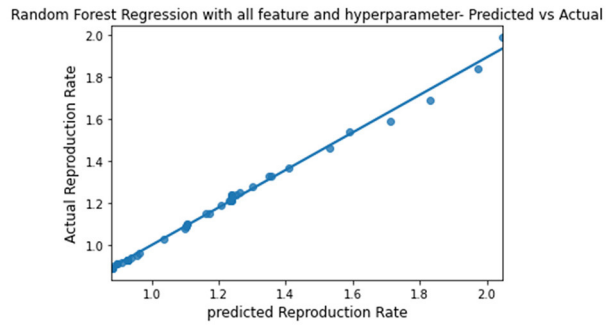


FIGURE 7 | Graph describing the predicted value vs. actual value of Random Forest.

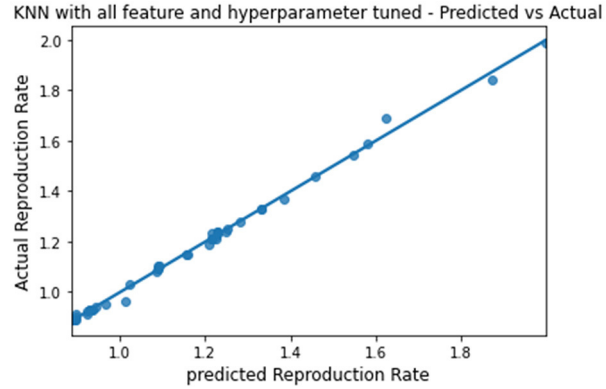


FIGURE 8 | Graph describing the predicted value vs. actual value of KNN.

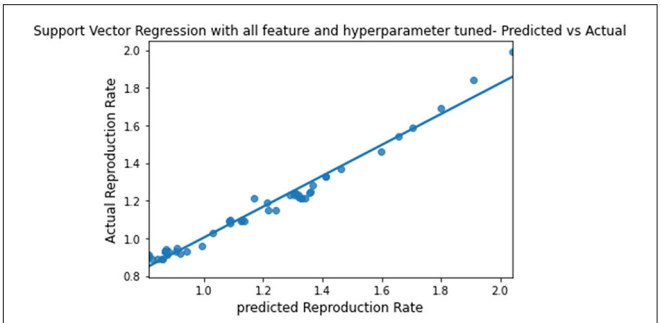


FIGURE 9 | Graph describing the predicted value vs. actual value of SVR.

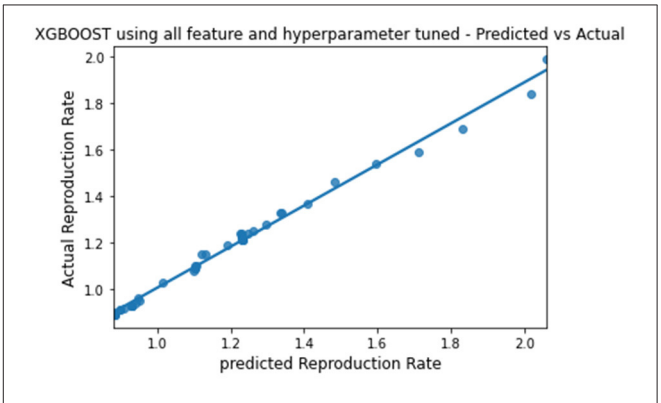


FIGURE 10 | Graph describing the predicted value vs. actual value of XGBoost.

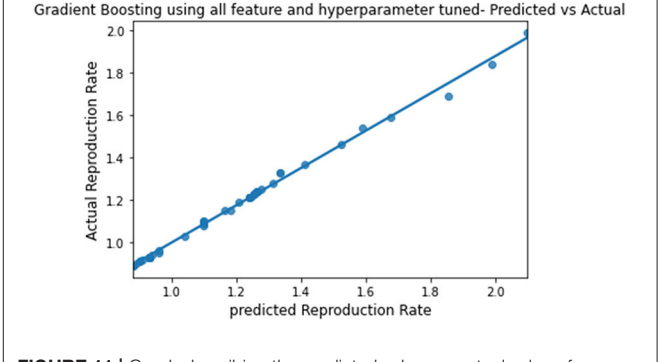


FIGURE 11 | Graph describing the predicted value vs. actual value of Gradient Boosting.

DISCUSSION

The major contributions of this paper are the study of features affecting the COVID-19 reproduction rate, as well as the investigation into the effects of feature selection and hyperparameter tuning on the prediction accuracy. Furthermore, prediction accuracy comparisons of the state-of-the-art regression techniques for COVID-19 reproduction rate have also been performed.

The selected features suggest that the total numbers of death and testing also influence the reproduction rate. Instead of depending only on the past value of the predictor variable as cited by Milind et al. (23), our work finds the crucial features affecting the predictor variable. Different regression techniques are used in the prediction and they are used to determine the

final reproduction rate. The effectiveness of feature selection in prediction has also been proven. Random forest has achieved the best performance in the accuracy comparison of the state-of-the-art techniques, as has already been proven by Chicco and Jurman (24). In the results obtained by the four experiments, the overall best values of MAE, MSE, RMSE, RAE, RRSE, and R-Squared were all obtained by the KNN approach. Therefore, KNN has obtained the best performance on average, followed by Random Forest and XGBOOST.

CONCLUSION AND FUTURE WORK

Predicting the reproduction rate is crucial, especially when a country has to take preventative measures to protect its citizens from a pandemic. Autoregressive models rely on and work with previous values to forecast future values. Non-linear machine learning regression algorithms have consistently produced the best prediction results in various applications, including the stock exchange, banking, and weather forecasting. Among the many factors involved in the spread of the COVID-19, the prominent factors are identified using Random Forest, Gradient Boosting, and XGBOOST in this work. Random Forest returned the highest importance score for Total_cases_per_million as 0.10196. For XGBOOST, the maximum score was 0.92185 for Total_case, and for Gradient Boosting, the top value of Total_deaths_per_million is 0.1183. Out of 16 features selected for investigation, seven features, namely, Total_cases, New_cases, Total_deaths, Total_cases_per_million, Total_deaths_per_million, Total_tests, Total_tests_per_thousand, and Positive_rate, are found to be prominent in reproduction rate prediction. Furthermore, this work investigated the reproduction rate prediction with non-linear machine learning regression techniques. The experiments were performed using Random Forest, Gradient Boosting, XGBOOST, KNN, and SVR, with and without feature selection and hyperparameter tuning. The results showed a decrease in the prediction error rate with hyperparameter tuning and with all of the features. Overall, the KNN algorithm had obtained the best performance. The study shows that Random Forest obtained the best performance with hyperparameter tuning and selected features. Individual regression techniques are applied in this study. However, the ensemble of regression techniques can be applied to obtain better performances. The regression algorithms obtained improved results with hyperparameter tuning and Gridsearch or Randomsearch methods. There is no remarkable difference in the prediction accuracy of algorithms with and without feature selection algorithms, so there is a need to find out the optimal features related to the reproduction rate.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

C-YC and KSr: conceptualization. C-YC: resources, project administration, and funding acquisition. JK and KSu: methodology and software. KSr, SM, and C-YC: validation. KSr, SM, and SC: writing—review and editing. JK: writing—original draft preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially funded by the Intelligent Recognition Industry Service Research Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. Grant number: N/A and the APC were funded by the aforementioned Project.

REFERENCES

1. Available online at: https://www.history.com/topics/middle-ages/pandemics-timeline (accessed May 20, 2021).
2. WHO (2020). Available online at: https://www.who.int/dg/speeches/detail/whodirector-general-s-opening-remarks-at-the-mission-briefing-oncovid-19-$-$12-march-2020 (accessed May 20, 2021).
3. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed May 21, 2021).
4. Wadhwa P, Aishwarya, Tripathi A, Singh P,Diwakar M, Kumar N. Predicting the time period of extension of lockdown due to increase in rate of COVID - 19 cases in India using machine learning. *Mater Today.* (2020) 37:2617–22. doi: 10.1016/j.matpr.2020.08.509
5. Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARSCoV-2 as compared with SARS-CoV-1. *N Engl J Med.* (2020) 382:1564–7. doi: 10.1056/NEJMc2004973
6. Gibson PG, Qin L, Puah SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. *Med. J. Australia.* (2020) 2:54–6. doi: 10.5694/mja2.50674
7. Bhattacharya SRK, Maddikunta PKR, Kaluri R, Singh S, Gadekallu TR, Alazab M, et al. A novel PCA-firefly based xgboost classification model for intrusion detection in networks using GPU. *Electronics.* (2020) 9:219. doi: 10.3390/electronics9020219
8. Luckner M, Topolski B, Mazurek M. Application of XGBoost algorithm in fingerprinting localisation task. In: *16th IFIP International Conference on Computer Information Systems and Industrial Management*. Bialystok: Springer (2017). p. 661–71. doi: 10.1007/978-3-319-59105-6_57
9. Vanden Driessche P. Reproduction numbers of infectious disease models. *Infect Dis Model.* (2017) 2:288–303. doi: 10.1016/j.idm.2017.06.002
10. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
11. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Krishnan SSR, Chowdhary CL, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc*. (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
12. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw Pract Exp*. (2020). doi: 10.1002/spe.2797. [Epub ahead of print].

13. Dhanamjayulu C, Nizhal UN, Maddikunta PK, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Processing*. (2021). doi: 10.1049/ipr2.12222. [Epub ahead of print].
14. Srinivasan K, Garg L, Chen B, Alaboudi AA, Jhanjhi NZ, Chang CT, et al. Expert system for stable power generation prediction in microbial fuel cell. *Intellig Automat Soft Comput*. (2021) 30:17–30. doi: 10.32604/iasc.2021.018380
15. Srinivasan K, Garg L, Datta D, Alaboudi AA, Jhanjhi NZ, Chang CT, et al. Performance comparison of deep cnn models for detecting driver's distraction. *Comput Mater Continua*. (2021) 68:4109–24. doi: 10.32604/cmc.2021.016736
16. Srinivasan K, Mahendran N, Vincent DR, Chang C-Y, Syed-Abdul S. Realizing an integrated multistage support vector machine model for augmented recognition of unipolar depression. *Electronics*. (2020) 9:647. doi: 10.3390/electronics9040647
17. Sundararajan K, Garg L, Srinivasan K, Bashir AK, Kaliappan J, Ganapathy GP, et al. A contemporary review on drought modeling using machine learning approaches. *CMES Comput Model Eng Sci*. (2021) 128:447–87. doi: 10.32604/cmes.2021.015528
18. Khosravi A, Chaman R, Rohani-Rasaf M, Zare F, Mehravaran S, Emamian MH. The basic reproduction number and prediction of the epidemic size of the novel coronavirus (COVID-19) in Shahroud, Iran. *Epidemiol Infect*. (2020) 148:e115, 1–7. doi: 10.1017/S0950268820001247
19. Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *Front Med*. (2020) 7:169. doi: 10.3389/fmed.2020.00169
20. Zivkovic M, Bacanin N, Venkatachalam K, Nayyar A, Djordjevic A, Strumberger I, et al. COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach. *Sustain Cities Soc*. (2021) 66:102669. doi: 10.1016/j.scs.2020.102669
21. Mojjada RK, Yadav A, Prabhu AV, Natarajan Y. Machine learning models for covid-19 future forecasting. *Mater. Today Proc*. (2020). doi: 10.1016/j.matpr.2020.10.962. [Epub ahead of print].
22. Farooq J, Bazaz A. A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India. *Alexandria Eng J*. (2020) 60:587–96. doi: 10.1016/j.aej.2020.09.037
23. Milind Y, Murukessan P, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos Solitons Fractals*. (2020) 139:110050. doi: 10.1016/j.chaos.2020.110050
24. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak*. (2020) 20:16. doi: 10.1186/s12911-020-1023-5
25. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li S-X, et al. Analysis of machine learning techniques for heart failure readmissions. *Circ Cardiovasc Qual Outcomes*. (2016) 9:629–40. doi: 10.1161/CIRCOUTCOMES.116.003039
26. Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos Solitons Fractals*. (2021) 142:110512. doi: 10.1016/j.chaos.2020.110512
27. Kuo CP, Fu JS. Evaluating the impact of mobility on COVID-19 pandemic with machine learning hybrid predictions. *Sci Total Environ*. (2021) 758:144151. doi: 10.1016/j.scitotenv.2020.144151
28. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. *Knowl Based Syst*. (2020) 25:106270. doi: 10.1016/j.knosys.2020.106270
29. Sujatha R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess*. (2020) 34:959–72. doi: 10.1007/s00477-020-01827-8
30. Yang Z, Zeng Z, Wang K, Wong S-S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*. (2020) 12:165–74. doi: 10.21037/jtd.2020.02.64
31. Xu Z, Huang G, Weinberger KQ, Zheng AX. Gradient boosted feature selection. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM (2014). p. 522–31.
32. Wang Y. A Xgboost risk model via feature selection and Bayesian hyper-parameter optimization. *arXiv:1901.08433* (2019).
33. Chintalapudi N, Battineni G, Sagaro GG, Amenta F. COVID-19 outbreak reproduction number estimations forecasting in Marche, Italy. *Int J Infect Dis*. (2020) 96:327–33. doi: 10.1016/j.ijid.2020.05.029
34. Locatelli I, Trächsel B, Rousson V. Estimating the basic reproduction number for COVID-19 in Western Europe. *PLoS ONE*. (2021) 16:e0248731. doi: 10.1371/journal.pone.0248731
35. Zhang Z, Trevino V, Hoseini S, Shahabuddin B, Smaranda B, Manivanna Zhang P, et al. Variable selection in logistic regression model with genetic algorithm. *Ann Transl Med*. (2018) 6:45. doi: 10.21037/atm.2018.01.15
36. Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 LSTM. *Sensors*. (2021) 21:2852. doi: 10.3390/s21082852
37. Panigrahi R, Borah S, Bhoi A, Ijaz M, Pramanik M, Kumar Y, et al. Consolidated decision tree-based intrusion detection system for binary and multiclass imbalanced datasets. *Mathematics*. (2021) 9:751. doi: 10.3390/math9070751
38. Panigrahi R, Borah S, Bhoi A, Ijaz M, Pramanik M, Jhaveri R, et al. Performance assessment of supervised classifiers for designing intrusion detection systems: a comprehensive review and recommendations for future research. *Mathematics*. (2021) 9:690. doi: 10.3390/math9060690
39. Tamang J. Dynamical properties of ion-acoustic waves in space plasma and its application to image encryption. *IEEE Access*. (2021) 9:18762–82. doi: 10.1109/ACCESS.2021.3054250
40. Chowdhary CL, Patel PV, Kathrotia KJ, Attique M, Perumal K, Ijaz MF. Analytical study of hybrid techniques for image encryption and decryption. *Sensors*. (2020) 20:5162. doi: 10.3390/s20185162
41. Gaur L, Singh G, Solanki A, Jhanjhi NZ, Bhatia U, Sharma S, et al. Disposition of youth in predicting sustainable development goals using the neuro-fuzzy and random forest algorithms. *Hum Cent Comput Inf Sci*. (2021) 11:24. doi: 10.22967/HCIS.2021.11.024
42. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA. (2016) p. 13–7. doi: 10.1145/2939672.2939785
43. Available online at: https://github.com/owid/covid-19-data/tree/master/public/data (accessed May 2, 2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kaliappan, Srinivasan, Mian Qaisar, Sundararajan, Chang and C. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fpubh.2021.737269

Blockchain-Secured Recommender System for Special Need Patients Using Deep Learning

Eric Appiah Mantey[1†], Conghua Zhou[1*†], Joseph Henry Anajemba[2†], Izuchukwu M. Okpalaoguchi[3†] and Onyeachonam Dominic-Mario Chiadika[4†]*

[1] School of Computer Science & Comm. Engineering, Jiangsu University, Zhenjiang, China, [2] Department of Communication Engineering, Hohai University, Changzhou, China, [3] Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria, [4] Electronic and Computer Engineering, Brunel University, London, United Kingdom

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Ebuka Ibeke, Robert Gordon University, United Kingdom
Gautam Srivastava, Brandon University, Canada

***Correspondence:**
Conghua Zhou
chzhou@ujs.edu.cn
Eric Appiah Mantey
ericphd@yahoo.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *06 July 2021*
Accepted: *03 August 2021*
Published: *20 September 2021*

Citation:
Mantey EA, Zhou C, Anajemba JH, Okpalaoguchi IM and Chiadika OD-M (2021) Blockchain-Secured Recommender System for Special Need Patients Using Deep Learning. Front. Public Health 9:737269. doi: 10.3389/fpubh.2021.737269

Recommender systems offer several advantages to hospital data management units and patients with special needs. These systems are more dependent on the extreme subtle hospital-patient data. Thus, disregarding the confidentiality of patients with special needs is not an option. In recent times, several proposed techniques failed to cryptographically guarantee the data privacy of the patients with special needs in the diet recommender systems (RSs) deployment. In order to tackle this pitfall, this paper incorporates a blockchain privacy system (BPS) into deep learning for a diet recommendation system for patients with special needs. Our proposed technique allows patients to get notifications about recommended treatments and medications based on their personalized data without revealing their confidential information. Additionally, the paper implemented machine and deep learning algorithms such as RNN, Logistic Regression, MLP, etc., on an Internet of Medical Things (IoMT) dataset acquired *via* the internet and hospitals that comprises the data of 50 patients with 13 features of various diseases and 1,000 products. The product section has a set of eight features. The IoMT data features were analyzed with BPS and further encoded prior to the application of deep and machine learning-based frameworks. The performance of the different machine and deep learning methods were carried out and the results verify that the long short-term memory (LSTM) technique is more effective than other schemes regarding prediction accuracy, precision, F1-measures, and recall in a secured blockchain privacy system. Results showed that 97.74% accuracy utilizing the LSTM deep learning model was attained. The precision of 98%, recall, and F1-measure of 99% each for the allowed class was also attained. For the disallowed class, the scores were 89, 73, and 80% for precision, recall, and F1-measure, respectively. The performance of our proposed BPS is subdivided into two categories: the secured communication channel of the recommendation system and an enhanced deep learning approach using health base medical dataset that spontaneously identifies what food a patient with special needs should have based on their disease and certain features including gender, weight, age, etc. The proposed system is outstanding as none of the earlier revised works of literature described a recommender system of this kind.

Keywords: blockchain privacy system, deep learning, machine learning, IOMT, artificial intelligence

INTRODUCTION

The idea of a secured recommendation system has been proposed in recent times due to the nature of diseases and dietary effects suffered by patients with unique conditions that inhibit their ability to receive routine diet recommendation care. This system which is incorporated with a blockchain privacy mechanism is advantageous to both the hospital data management unit and the patients with special needs in terms of privacy violation protection, scandals, and longevity of the patients. In (1), the authors gave a comprehensive explanation, including their history and basic concepts of networks and how they could be applied in the field of pharmaceutical sciences as an option instead of just traditional methodology. The authors of (2) discussed the various ways Artificial Neural Networks (ANNs) can be applied to resolve challenges in the pharmaceutical field. Similarly, in (3), the authors discussed how ANNs can be used in diverse ways in medicine such as in the delivery of drugs, classification of cancer, research in pharmaceutics and others. The authors of (4) proposed an algorithm to solve the issue of classification, analysis, and summarization of document analytics. Experimental results showed the efficiency of the algorithm regarding its precision and implementation time. In (5), the authors proposed a solution based on deep learning for medical datasets which identifies what kind of food a patient should be fed based on factors like the nature of diseases, gender, age, and weight, among others. Their framework made use of machine learning and deep learning algorithms. Experimental results showed its efficiency and accuracy.

The motivation of this study is to enhance, secure, and analyze the performance of an expansive nutritional theory into the Internet of Medical Things (IoMT) using a blockchain privacy system and deep learning algorithms.

The contributions of this study are stated below:

- To propose and incorporate a reliable blockchain privacy system (BPS) for a secured diet recommendation system for patients with special needs.
- To investigate and perform analysis of machine learning, deep learning system previously used on IoMT such as Naive Bayes and logistic regression, recurrent neural network (RNN), gated recurrent unit (GRU), and long short-term memory (LSTM).
- To design an all-inclusive diet recommendation model which can be applied for patient special needs products and disease specifications.
- To analyze the behavior of our Enhanced Artificial Intelligence (AI) and deep learning mechanisms and how it is administered.
- The result showcased the Enhanced Machine Learning and deep learning for patients with special needs and for treating varied patient diseases with varied recommender evidence that is secured using a blockchain privacy system.

Structure

The remaining part of the study laid out on this paper is organized as follows. Several literature are revised and detailed in the related works section. Introduced in section System Model are the methods and materials of the system, including its execution utilizing AI. Summarized in section Experiments and Results are the findings of the conducted experiment in this study. Finally, the conclusions of the study are discussed in section Conclusion and Future Work.

RELATED WORKS

Blockchain is one of the most recent subjects currently understudied and has been recently incorporated into several societal, industrial and academic scenarios (6). One of the areas that have extensively incorporated the BPS is the hospital and patient recommender systems. The authors in (7) proposed research for customizing a service for the recommendation of diet for its customers which prevents and manages heart diseases in the area of healthcare. It provides this service by putting into consideration factors like family history of illnesses, preferences of food, vital signs, etc., of the customers with the issue of heart diseases. This service can aid customers to change their living habits to healthier ones. The study conducted by the authors of (8) proposed a framework for nourishment recommendation where input is retrieved from children and analyzed, which results in output that provides for a better and healthier diet plan. Its goal is to provide kids from the age of eight to their early teens with a healthier eating lifestyle which varies according to their age, gender, and other related factors. The study of the authors in (9) suggested a platform that smartly plans the meals of users according to their health conditions. This platform makes use of machine learning algorithms to function. The authors of (10) studied the impact of various factors which influence the adoption and use of a user of online health facilities, specifically in Ukraine and China. The authors in (11) proposed a framework that recommends healthy diet plans to patients that have hypertension. It utilizes machine learning algorithms and considers factors such as allergies, food preferences, age, and blood pressure, among others. The study conducted by the authors of (12) discussed how blockchain technology in conjunction with the Internet of Things (IoT) could be used to monitor the supply chain efficiently. The \ study of the authors in (13) proposed a system that recommends healthier diets specifically for American Indians with diabetes, by studying the profile of the user with clinical regulations and guidelines. The system makes customized recommendations based on various factors for the AI patient. The authors of (14) used a neural network deep learning method to explain how big data analytics could be utilized in the execution of an efficient system for health recommendation, and demonstrates how the industry of healthcare could evolve from a customary system to a more customized one in the online health environment.

The authors in (15) examined the ways for calculating and measuring the similarities of users in health websites. The recommendation of similar users to customers aids in the support search, whether emotional or not in a more efficient way. The study conducted by the authors of (16) presented an overview of the literature that examines how machine learning algorithms are used in recommender systems and also pinpoints

opportunities for research in the software engineering field. It concluded that the algorithms are used because they are not complex. The study of the authors in (17) proposed a method for recognizing parameters that influence processes and suggests a framework for machines by executing a device that can collect a large amount of data and integrate them into the Cloud for more studies. The authors of (18) identified various ways big data analytics could be utilized and gave its benefits regarding IT framework in strategic areas. They also suggested tactics for organizations concerned with healthcare to adopt big data analytics. The authors in (19) proposed a recommendation of threads to health communities' online users by making use of diverse healthcare information network mining. The study conducted by the authors of (20) proposed a scheme to improve the efficacy of recommender systems and preserve privacy, as long as the data collected are examined efficiently. The study of the author in (21) provided an exhaustive review of all research regarding and related to recommender systems that are based on deep learning technology. The authors of (22) proposed a model that is based on fog deep learning that acquires data from people and forecasts their fitness statistics and abnormalities making use of a model that is based on a neural network that is capable of handling diverse and large amounts of data.

The authors in (23) proposed a recommender system that delivers accommodative nutrition information to extend and improve the lives of people with diet-related illnesses, as well as healthy people. The system recommends nutrition according to the health profile of the user. The study conducted by the authors of (24) proposed an architecture that collects and collates data from different primary indicators of performance in healthcare communities, as well as predicts possible values of these principal performance indicators. The study of the authors in (25) proposed a system that utilizes customized features concerning nutrition that aids users in the change of their attitude toward healthier food choices and diets. The authors of (26) conducted a review to determine if dieted and nutritional products are issues often left out in shortages research. The authors in (27) proposed a framework for the recommendation of everyday meals while also managing preference-inclusive and nutritional information. The study conducted by the authors of (28) proposed a recommender system making use of analysis of food clustering for patients with diabetes. The study of the authors in (29) reviewed numerous existing works of literature and discovered lacunas in the existing systems of diet management. The authors of (30) in their study concentrated on developing a recommender system that integrates Artificial Intelligence methods and creates a base of knowledge following the appropriate and lawful guidelines and regulations concerning diabetes. The recommended menu is according to the preferences and conditions of the patients.

The authors in (31) presented a recommender system that is based on the pathology report of the user. This system utilizes the ant colony algorithm to create a menu and suggests appropriate food in line with the values of the pathology report of the user. The study conducted by the authors of (32–34) discussed error sources regarding nutrition, as well as possible solutions to tackle those challenges for future systems. They also discussed the incorporation of nutritional propositions into information systems. The study of the authors in (35) proposed a management system to aid in the preparation of a proper and healthy diet for children. The authors of (36) studied the shifts in obesity and its various kinds from the richer to the less fortunate across an era of economic growth in a developing country (Malaysia). They concluded that as the country develops, the risk of obesity among the less fortunate should be expected. The authors in (37) proposed a customized recommender system fashioned to meet the nutritional restrictions, preferences, and expectations of the user. The conducted by the authors of (38) recommended a diet plan to patients with jaundice using a hierarchy process that considers vital nutrients to determine the best choice for meals for all times of the day. The study of the authors in (39) discussed the uses and benefits of deep learning methods and ANNs in the advancement of the medical field in areas such as electroencephalography and the collation of physiological data. On the other hand, the authors in (40) presented a review of recommendation methods for people and groups to eat healthier foods. They also studied the existing recommender systems and challenges regarding research of recommendation technology. The study of the authors in (41) presented a Privacy-Preserving and Secure Framework (PPSF) for IoT-driven smart cities. The proposed PPSF is subject to two key schemes: a two-level privacy mechanism and an intrusion detection mechanism. Finally, the potentials of blockchain-enabled edge-of-Things (BEoT) by Prabadevi et al. in (42) was designed to provide security and services which comprises trust management, attack detection, data privacy preservation and access authentication.

SYSTEM MODEL

The methodology used in this study is explained in the following discussions. The objective of this study is for the recommendation of diet to various patients with special needs with a BPS utilizing machine and deep learning classifiers for the health-based medical dataset which will spontaneously identify what kind of food a patient with special needs should have, based on their disease and other factors such as weight, gender, and age, among others. For this reason, we have incorporated a BPS utilizing various deep and machine learning classifiers. A random forest classifier was utilized to know what feature has a greater effect on the dataset.

Dataset

The dataset utilized in this study comprises about a thousand products and 50 patients collated using IoT and cloud methods. The products experimented on various disease patients. The dataset possesses over 13 features and has stored in it almost 17,000 records. The features of the products are listed in **Table 1**, the features of the patient are listed in **Table 2**, while **Table 3** is the Training accuracy of models after BPS.

Data Preservation and Protection

The proposed BPS can be used by patients to store and preserve their personal data which are applicable to recommender systems. Example of such data includes diagnosed

TABLE 1 | Number of features in the product.

S/No.	Product features	Feature type
1	Product barcode	Numeric
2	Product calories	Numeric
3	Product protein	Numeric
4	Product Fat	Numeric
5	Product sodium	Numeric
6	Product carbohydrate	Numeric
7	Product fiber	Numeric
8	Product cholesterol	Numeric

TABLE 2 | Features of patients.

S/No.	Patients features	Feature type
1	Patient number	Numeric
2	Patient age	Numeric
3	Patient gender	Categorical
4	Patient weight	Numeric
5	Patient disease	Categorical
6	Patient calories	Numeric
7	Patient protein	Numeric
8	Patient fat	Numeric
9	Patient sodium	Numeric
10	Patient carbohydrate	Numeric
11	Patient fiber	Numeric
12	Patient cholesterol	Numeric
13	Target class	Categorical

TABLE 3 | Training accuracy of models after BPS.

Model name	Average accuracy %
MLP	92.91
RNN	94.5
GRU	95.29
LSTM	96.5
Naïve Bayes	92.81
Logistic Regression	92.79

illness, treatment status, demographic attributes, prescribed medications, and response performance, among others. The BPS stores and fully encrypts this set of data, such that, accessing them without permission is almost impossible. Other related data, such as health history, diets of the patients, and more, which illustrates the actual treatment performance that is vital for the recommender systems are classically gathered by the hospital database system (HDS). These user-sensitive data are transmitted and stored in the BPS instead of the database of the hospitals, as illustrated in **Figure 1**, in order to enhance privacy by preventing attacks of the eavesdroppers, hacker attacks, and other intruding vulnerabilities. On the other hand, the BPS grants adequate permission to the patients to access these data in a very secured and trusted format. In case of a recommendation, prescription, check-ups, or emergencies, the BPS also sends a prompt notification to the registered smart device of the patient. This ensures adequate transparency about the data collection events of the hospitals for the patients (users).

Cooperative Filtering

One of the major advantages of the BPS over other existing privacy technology is its ability to perform efficient data computations while keeping all input data confidential. As illustrated in **Figure 2**, the hospital is notified each time the patient (user) attempts to obtain a recommendation for their confidential medical details. Because the hospital runs a multi-user system with several of its patients, their individual data are employed as input variables for the cooperative filtering algorithm of the recommender systems to sort and locate the best recommendations for each enquiring patient. Meanwhile, the hospital database manager can only view basic information of the patient as the rest of the sensitive data and computations are performed at the BPS. In conclusion, the patient gets a notification alert and receives secure permission to access the subsequent recommendations.

Data Processing

Data Normalization

Functions of data cleaning are carried out to get rid of noise and regularize the features after choosing a dataset. The aim of regularization is for the dataset to be scaled into a single range. This is done because the dataset has varied scale values. They can be one-digit, two-digit, or three-digit values so every value is brought into one scale so that the machine learning models perform more effectively. For this reason, we carried out min-max normalization. In this study, min-max scaling normalizes values within the scope of [0, 1] and is illustrated in Equation (1) below.

$$N_i = \frac{c_i - min(c_i)}{max(c_i) - min(c_i)} \tag{1}$$

where $y = (C_1, C_2 ... C_n)$ are the number of features. C_i is the feature to be normalized and N_i stands for the features that have been formalized. As a result of doing this, all the features now have identical weights, and they are all in a single scope.

Data Encoding

While this research was carried out, values that were duplicated and inconsistent were eliminated from the dataset before data encoding was performed. Subsequently, the nominal features are transformed into arithmetic values. The objective of this is to ensure that the backend functions of the machine learning models are performed using arithmetic values prior to their implementation. In this study, non-arithmetic data was transformed to arithmetic data prior to the performance of data encoding. Machine learning (ML) algorithms backend computations were also carried out on arithmetic values before data was passed to the suggested model.

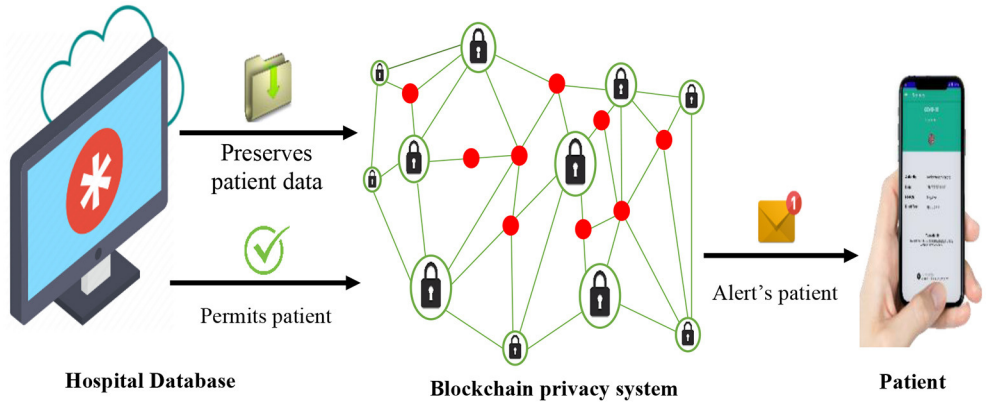


FIGURE 1 | BPS-based data storage and patient alert preservation system.

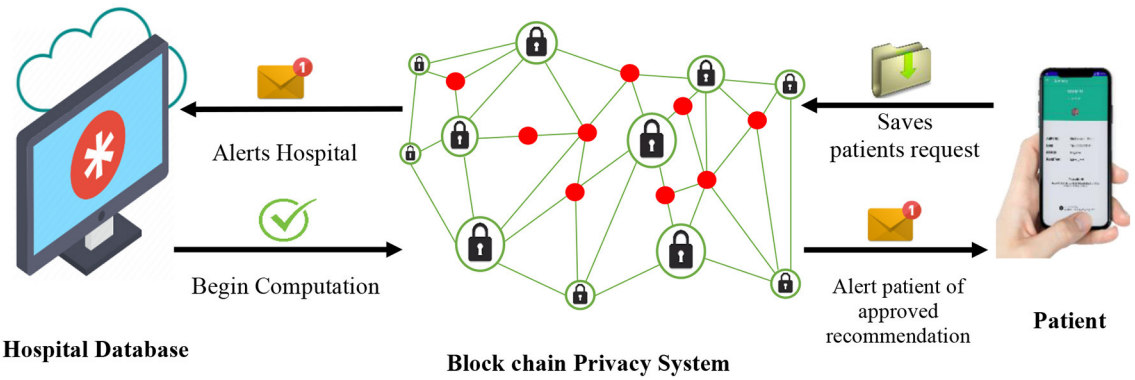


FIGURE 2 | BPS-based data storage, hospital management and preservation system.

Optimal Feature Visualization

A Product calorie has an important percentage of 48% in the dataset. Product fat has a significance percentage of 12%. Product protein and carbohydrates have significance percentages of 8%. Product sodium has an important percentage of 6%. The user number has an important percentage of 5%. Furthermore, product fiber, user fat, and protein have significance percentages of 2% each. Age, user weight and calories, disease, product barcode, and user carbohydrate have important percentages of 1% each in the dataset. Random Forest is a combination of decision trees. For the prediction to be correct, it integrates all the decision trees and gives more precise results. It is used not just for regression and classification, but also in the application of the best features in the dataset. Its best attribute is that regression and classification can be performed using it. When classifying Random Forest, the majority vote was utilized in the prediction of the target. However, for the evaluation of regression, Random Forest takes the average value of all the decision trees and predicts it as a threshold as stipulated for each node. Subsequently, there is a performance of splitting based on that threshold. The threshold is stipulated by computing gain-index and entropy. Their various equations are shown below.

$$Entropy: K(b_1, b_2, ...b_s) = \sum_{i=1}^{s} (b_1 \, log(\frac{1}{b_1}) \quad (2)$$

where b_1, b_2...b_s stands for the probabilities of the class labels.

$$Gain(R, S) = K(R) \sum_{i=1}^{S} x \, (R_1) \, K(R_1) \quad (3)$$

Deep Learning Classifiers

Multilayer Perception

There exist various kinds of neural networks created and developed previously. All the existing ANNs may be distinguished by their processing unit (PE) transfer functions. Their techniques of learning and by the link equations. Processing unit is a basic constituent of ANN. It collects numerous signals that are weighted from other existing processing units. The framework of forwarding ANN is like this. The first layer has its input units. There are hidden layers in the middle. Finally, the last layer comprises output units. The function of the input unit is the provision of data from

outside sources. Next, the data that is collected is transported to concealed layers where it is amplified with weights. Finally, it is passed to the output layer to initiate the final signals. The classification capacity of ANN is based completely on concealed layers. These concealed layers are linked by the synapses with the neighbor's layers. If there are *p* input data points (b_1, b_2... b_p), we call it *p* number of features in the data. In the ANN framework, all the features are amplified with weights (e_1, e_2...e_p) and then added as is stated in Equation (4) below.

$$E.B = e_1b_1 + e_2b_2 + ... + e_pb_p = \sum_{1=1}^{p} e_ib_i \quad (4)$$

where *"p"* represents the total amount of features present in the dataset given as input B to the input layer. While *"e"* stands for the weights of all the features amplified with its weight. It is referred to as the dot product too. The addition of the bias function is done next in the dot product function. It will lead to Equation (5) stated below.

$$a = \sum_{1=1}^{p} e_ib_i + bias \quad (5)$$

In Equation (5), *a* denotes the activation function f (a). In this manner, we will obtain the output for the first neuron and the first hidden layer. This procedure will be repeated until the last weight and the last input is seen.

Recurrent Neural Network

Recurrent Neural Network is a kind of ANN where node-to-node links create a graph administered along with a corporeal order which makes provision for the display of dynamic behavior. It is a type of new ANN that has memory cycles that are directed.

Long Short-Term Memory

This is a framework that carries out a memory augmentation for the RNN. In this study, 3 layers of LSTM with a batch size of 32, as well as the sigmoid function is used to activate it. The Adam function is utilized for optimization. Binary cross-entropy is used to compute function loss.

Gated Recurrent Units

The principle of the GRU is more current than the LSTM. Basically, it functions more effectively. It is quicker in the training of models as opposed to the LSTM. The model can be modified and manipulated easily. However, LSTM performs better than the GRU if a longer memory is needed. Eventually, the comparison of performance depends on the kind of dataset being used. The frameworks LSTM and GRU are somewhat similar, but there are some important differences stated below:

- The GRU comprises two gates, but the LSTM has three.
- Gated recurrent units do not have internal memory that is opposed to the detectable hidden form. The output gate which is embedded in LSTMs is also not present in the GRUs.
- There is no application of second non-linearity when calculating GRU output as opposed to the LSTMs.

Machine Learning Classifiers

Logistic Regression

This is a popular classification algorithm utilized in machine learning. Basically, a bipartite result is acquired with it. Its purpose is to find a link connecting the possibility of a particular outcome and its features. A logit (odds) function is utilized in this algorithm. In Equation (6), the algorithm is described as seen below:

$$\log\left(\frac{r(b)}{1-r(b)}\right) = \beta_o + J_{1b} \quad (6)$$

In the equation above, the logit function is $\log\left(\frac{r(b)}{1-r(b)}\right)$ and the odd function is $\left(\frac{\mathbf{r(b)}}{\mathbf{1-r(b)}}\right)$. The odds show the probability proportion of insertion of the feature to the likelihood of failure or lack of features. Output is normally done in this algorithm subsequent to the mapping of inputs to log-odds in a straight blend. Now, the contrary of the function stated above would be;

$$R(B) = \frac{1^{\beta_o + \beta_{1b}}}{1 + 1^{\beta_o + \beta_{1b}}} \quad (7)$$

Equation (7) is referred to as a sigmoid function which is a mathematical function that has an "S"—shaped curve characteristic, which converts the values between the range 0 and 1. As long as the probability value is initiated within the range of $0 < R < 1$, then we chose the variables in the logarithm in a manner to expand the likelihood of scrutinizing sample values.

Naive Bayes

This is a collation of algorithms that have a joint where all the pairs of the features are unconstrained. Two hypotheses are considered in the algorithm which are as follows:

1. All the functions are separate.
2. The texts need to be translated into arithmetic values in the instance of features in text format. Bayes' Theorem is shown in Equation (8) below:

$$R\frac{G}{H} = \frac{R\left(\frac{G}{H}\right)R(G)}{R(H)} \quad (8)$$

where G and H events, R (G) is the prior probability $R\left(\frac{G}{H}\right) = H's$ posterior probability. Now, Bayes' theorem can be implemented by Equation (9) as seen below:

$$R\frac{v}{B} = \frac{R\left(\frac{B}{v}\right)R(v)}{R(B)} \quad (9)$$

where B represents the dependent feature vector and v is the class variable, B is of size n so, B = (B_1, B_2, B_3 ..., Bn). For evidence to be split into unconstrained categories for events G and H: R (G, H) =R (G) R (H). The results then become:

$$R\left(\frac{v}{B_1 \ldots n}\right) = \frac{R\left(\frac{B_1}{v}\right)R\left(\frac{b_2}{v}\right)\ldots\ldots R(\frac{b_n}{v})}{R(B_1)R(B_2)\ldots\ldots R(B_n)} \quad (10)$$

It can be expressed as Equation (11) below:

$$R\left(\frac{v}{b_1 \ldots n}\right) = \frac{R(v)\Pi_{i=1}^{n} R\left(\frac{B_i}{v}\right)}{R(B_1)R(B_2)\ldots\ldots R(B_n)} \quad (11)$$

Because the denominator is continuous for input, so in Equation (12) below:

$$R\left(\frac{v}{b_1 \ldots n}\right) = \infty R(v)\Pi_{i=1}^{n} R\left(\frac{B_i}{v}\right) \quad (12)$$

For a classifier model to be produced, the input probabilities for v must be determined and the output with the highest probability has to be taken. Hence, in Equation (13) below:

$$v = argmax R(v)\Pi_{i=1}^{n} R\left(\frac{B_i}{v}\right) \quad (13)$$

Finally, what is left for calculation is $R\left(\frac{b_i}{v}\right)$ which is the initial probability and R (v) which represents the contrast of the logit function.

Evaluation Metrics

Various metrics are utilized in the evaluation of the performance of our suggested model. These are stated below:

$$Accuracy = \frac{A_z + A_n}{A_z + C_z + C_n + A_n} \quad (14)$$

The objective of precision is to examine the True Positive (AZ) units in connection to False Positive (CZ) units.

$$Precision = \frac{A_z}{A_z + C_z} \quad (15)$$

The objective of recall is to examine True Positive (AZ) units in connection to False Negative (CN) units that are not classified. The arithmetic arrangement of recall is stated in Equation (16) below:

$$Recall = \frac{A_z}{A_z + C_n} \quad (16)$$

Sometimes, assessment of performance may not be very efficient with recall and accuracy. For example, if a mining algorithm has high precision but low recall, then another algorithm is needed. Then comes the question of which algorithm is more effective. This challenge is solved utilizing F1-measure that gives a mean recall and precision. F1-measure can be calculated as follows:

$$Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (17)$$

Software Analysis

A Colab software installed on a Core-i7 computer system is employed in performing the Deep Learning experiments. The system is fully equipped with 16GB RAM, 4 CPUs, a 1.7 GHz processor, and about 20GB obtained from Google Colab laboratory. The overall dataset was partitioned into three segments which includes training segment, cross-validation segment, and testing segments. Constituting 80% of the dataset was utilized in the research as the training set, 20% was used as the testing set, while *K*-Fold Cross-validation was employed and is applicable to both the training and testing sets.

TABLE 4 | Performance of training score accuracy for selected classifiers.

Classifier	Percentage accuracy (%)
LSTM	95.5
RNN	92.3
MLP	86.5
Logistic Regression	88.5
Naive Bayes	87.2
GRU	94.6

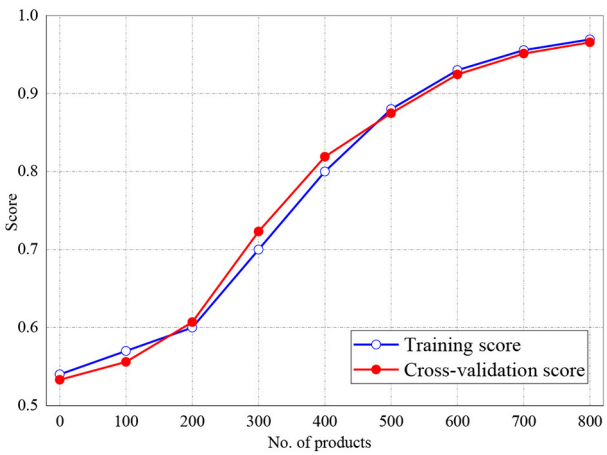


FIGURE 3 | Result for Training and Cross-Validation of Naive Bayes classifier.

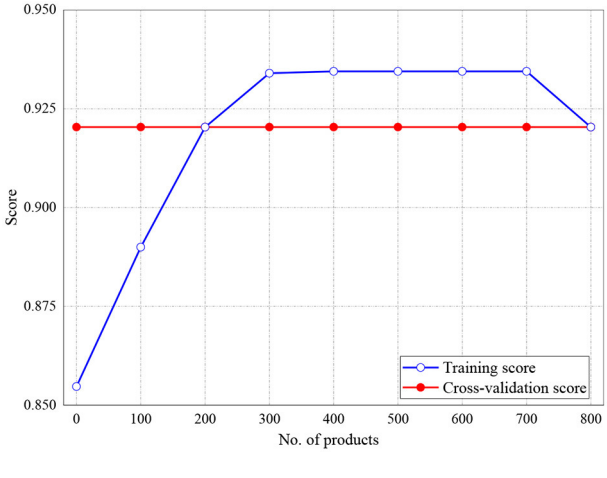


FIGURE 4 | Result for Training and Cross-Validation Scores of Logistic Regression classifier.

The study employed both machine and deep learning classifiers in this research and their training accuracies are presented in **Table 4**. Using a 32-batch sized 3-layer LSTM with sigmoid activation function, it is established that the LSTM classifier ranked maximum with 95.45% training accuracy, while MLP ranked as the least at 86.5%. The training accuracies for the rest classifiers which are GRU, RNN, Logistic regression [LR], and Naive Bayes were at 94.6, 92.3, 88.5, and 87.2% training accuracies, respectively.

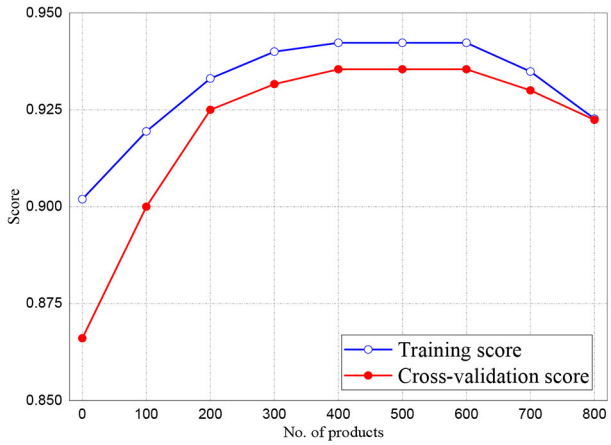


FIGURE 5 | Result for Training and Cross-Validation of MLP classifier.

TABLE 5 | Performance of testing score accuracy for selected classifiers.

Classifier	Percentage accuracy (%)
LSTM	99.5
RNN	95.5
MLP	90.3
Logistic Regression	90.8
Naive Bayes	95.6
GRU	98.8

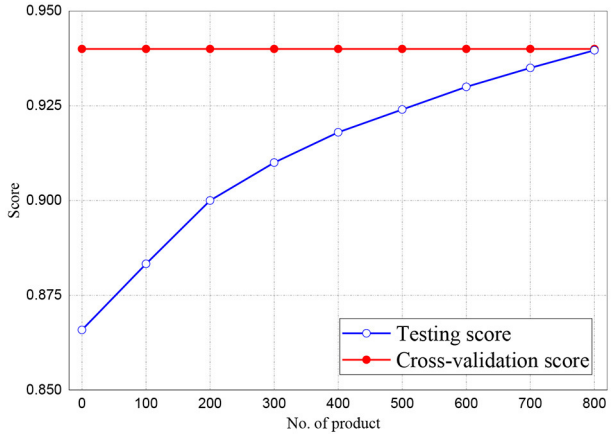


FIGURE 6 | Result for Testing and Cross-Validation of Naive Bayes classifier.

EXPERIMENTS AND RESULTS

The training and validation scores of Naive Bayes, LR, and MLP classifiers are presented in **Figures 2**, **4**, **5**, respectively. For all the figures, the training scores are denoted with a blue line/curve while the red curve represents the scores of cross-validations. For Naive Bayes in **Figure 3**, it is observed that the training and cross-validation scores merged at 87.2%. The cross-validation score of the logistic regression in **Figure 4** remained stable and linear while the training score increases up to 93.8%, stabilizes, and further decreases. The validation and training scores of the MLP classifier is illustrated in **Figure 5**. The experiment shows that both the training and validation scores of the MLP classifier training scores increase at some point and ultimately decreased.

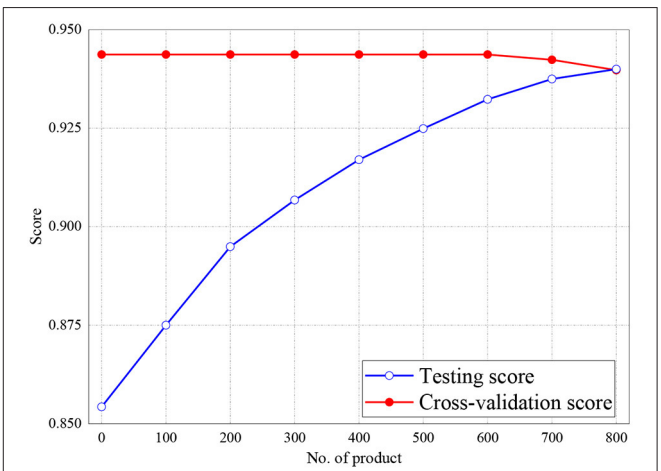


FIGURE 7 | Result for Testing and Cross-Validation of Logistic Regression.

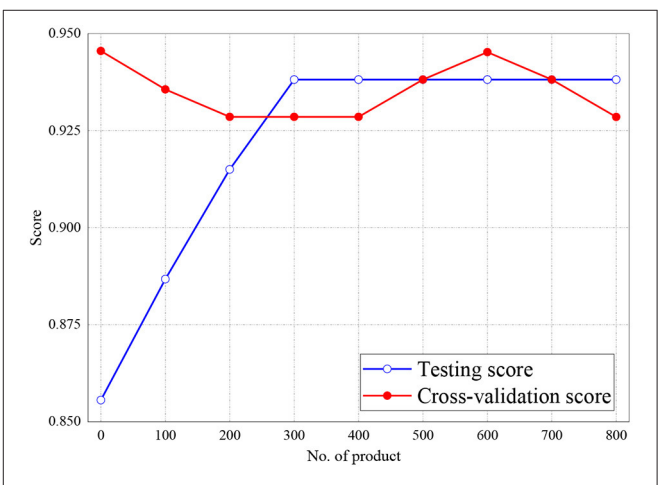


FIGURE 8 | Result of Testing and Cross-Validation for MLP classifier.

The test accuracies of the MLP classifiers are presented in **Table 5**. The result shows that the LSTM classifier realized a maximum of 99.5% test accuracy, while the MLP classifier achieved the least score of 90.3% test accuracy. The rest classifiers which are GRU, Naive Bayes, RNN, and LR all attain 98.8, 95.6, 95.5, and 90.8% testing accuracies, respectively.

Figures 6, **7** represent the testing accuracies and cross-validation scores of both Naive Bayes and LR models, respectively, it is observed that the testing accuracies and validation scores for both models merged at 94 and 94.28%, respectively. Similarly, **Figure 8** shows that the validation and testing scores for MLP merged at 93.81 and 92.85%, respectively.

The study further investigated the accuracies and loss performance of training scores against the testing scores for all the models. The results of these investigations are presented in **Figures 9–11**. The red curves denote the training scores, while the blue curves denote the testing scores for both the accuracy and loss performances. **Figure 9A** represents the performance accuracy of the training and testing scores for the GRU classifier. It is observed that the performance started from 87.5% and remained stable until after about 45 epochs. It further witnessed instabilities till it achieved an apex of 93% at 96.6 epoch. Similarly, the blue curve which represents the testing score started and remains stable at 85.5%, witnessed several instabilities but attained its apex of 90.2% at 100 epochs.

On the other hand, the loss performance of the training and testing scores for the GRU classifier is represented in **Figure 9B**. The figure indicates that the loss performance of the training

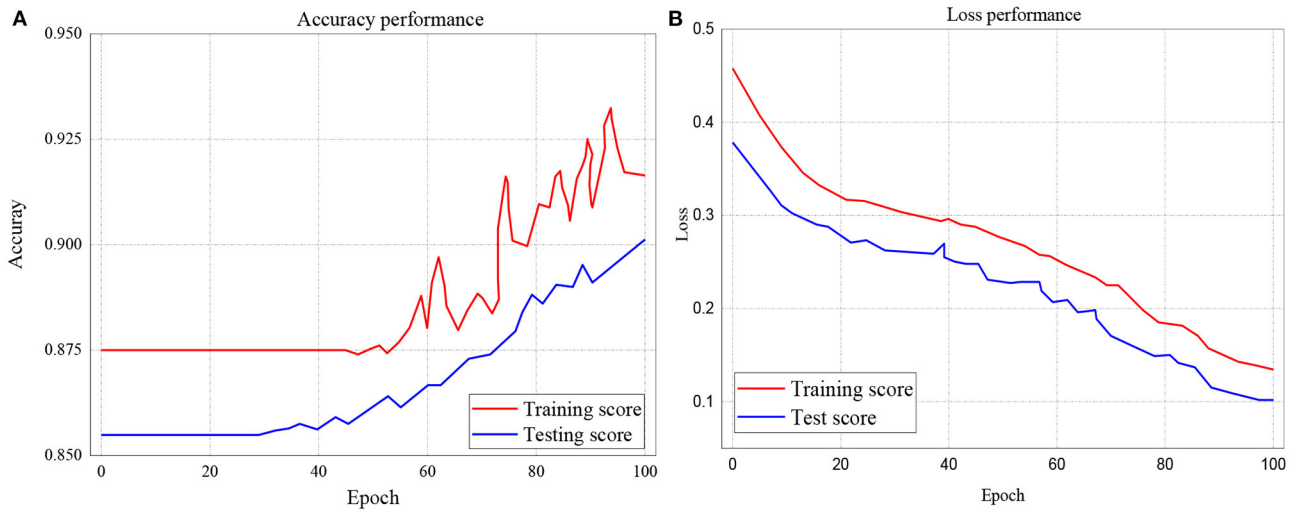


FIGURE 9 | (A,B) Performance accuracy and loss of Training and Testing Scores for GRU classifier.

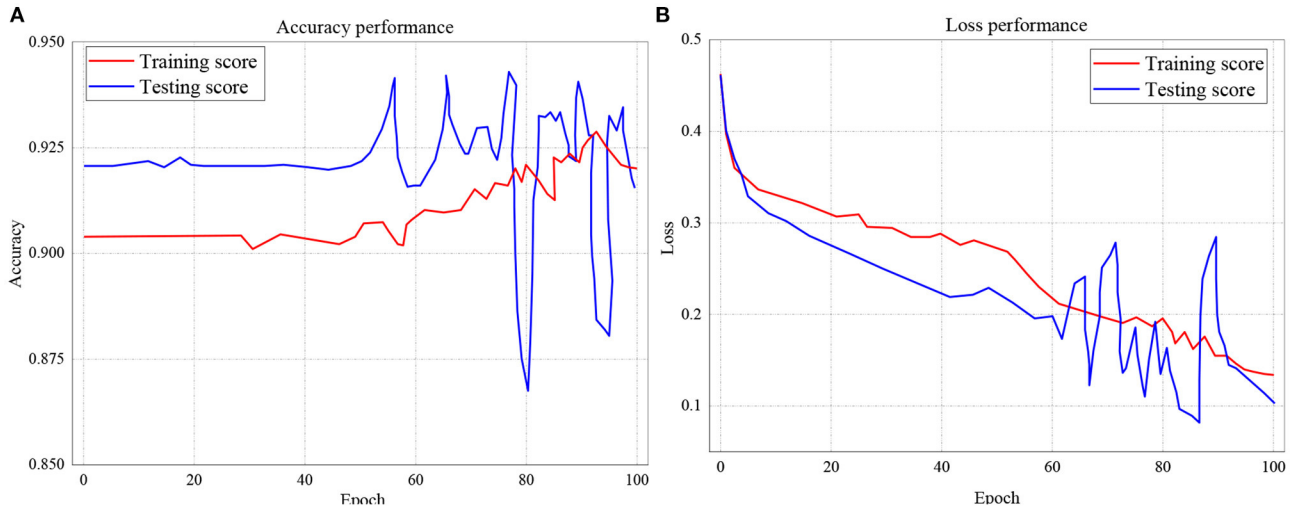


FIGURE 10 | (A,B) Performance accuracy and loss of Training and Testing Scores for LSTM.

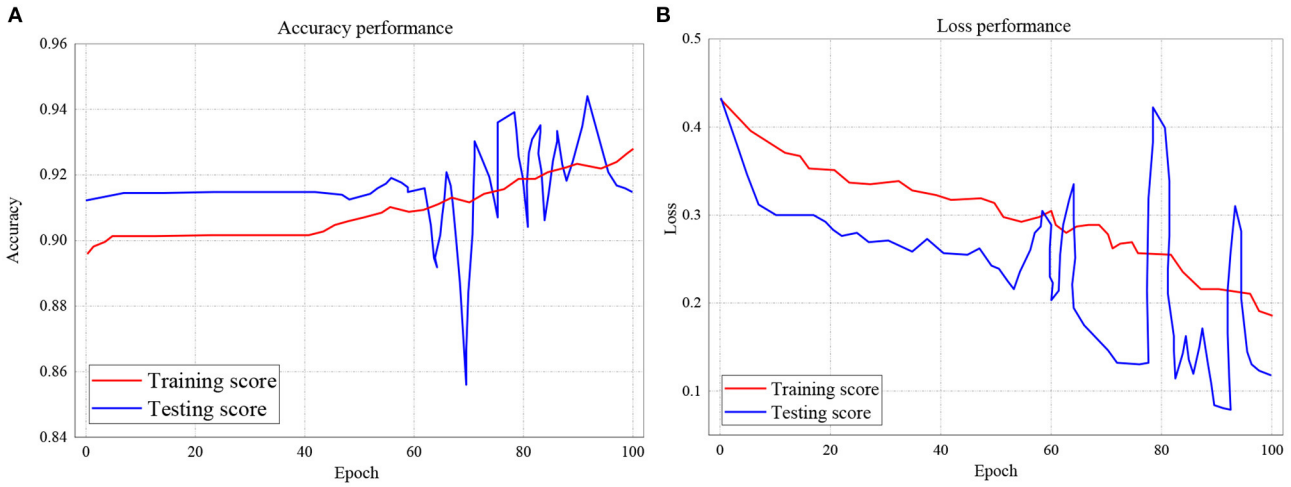


FIGURE 11 | (A,B) Performance accuracy and loss of Training and Testing Scores for RNN.

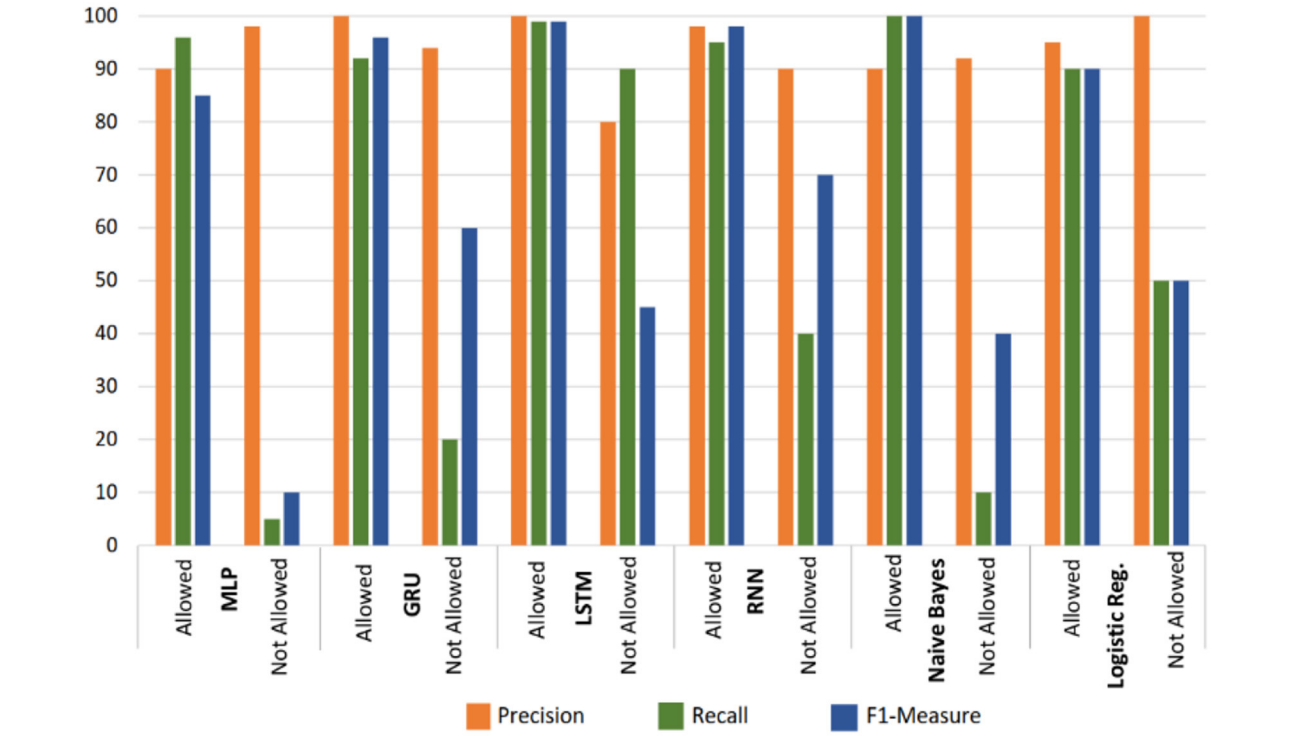


FIGURE 12 | Result of the Classification for Deep and Machine Learning classifiers.

score began from 0.45 and decreased to 0.14. In the same way, testing loss performance commenced from 0.38 down to 0.102.

The training and testing scores accuracy for LSTM is presented in **Figure 10A**, while the loss performance for the same model is illustrated in **Figure 10B**. Considering **Figure 10A**, the experiment shows that for accuracy, the training score started from 92.1%, with the least performance of 86.8% and peak performance of 94.3%. Likewise, for testing score accuracy, the performance began from 90.4% to a peak of 92.8% with the least performance of 90.1%. On the other hand, the training and testing loss performance of the LSTM classifier is illustrated in **Figure 10B**. The performance indicates that the training loss score commenced from 0.46 and declines to 0.14 after 100 epochs. The testing loss performance shows that the

reading also started from 0.46 but decreased down to 0.1 after 100 epochs.

In **Figure 11A**, the testing and training scores accuracy for the RNN model is presented, while the loss performance for the same model is illustrated in **Figure 11B**. With respect to **Figure 11A**, the experiment shows that for accuracy, the training score started from 89.5% and also records as the least performance with a peak performance of 92.9% after 100 epochs. Also, for testing score accuracy, the performance began from 91.3% to a peak of 91.5% with the least performance of 85.8%. On the other hand, the training and testing loss performance of the RNN classifier is illustrated in **Figure 11B**. The performance indicates that the training loss score commenced from 0.435 and declines to 0.185 after 100 epochs. The testing loss performance shows that the reading also started from 0.435 but decreased down to 0.12 after 100 epochs.

The classification further recorded the deep learning and machine learning classifiers which are presented in **Figure 12**. The result indicates that the LSTM classifier outperforms the rest of the tested classifiers with reference to recall, F1-measure, and precision. The experiment is divided into allowed and not allowed classes. The performance of the LSTM classifier for the allowed class is measured at 99, 100, and 99% for recall, precision and F1-measure accordingly. Meanwhile, the disallowed class performance of the LSTM classifier measure at 90, 80, and 45%, for recall, precision, and F1-measure, respectively. The rest of the experimented models performs well in terms of the allowed class. However, they did not perform well with the disallowed class.

CONCLUSION AND FUTURE WORK

The analysis of collected automated medical data has been established to have the potential in providing adequate recommendations for the improvement and management of patients with special needs in the hospitals through different algorithm developments and knowledge discovery. However, several medical facilities and special care patients are yet to grasp the usage and intent of secured recommender systems. One of the reasons for not embracing the systems is due to fear of data Privacy Breach. This research proposes a secure deep learning-based recommender system that estimates and issues basic treatment and diet recommendations to patients with special needs without revealing their sensitive health details. The system automatically issues exact recommendations to ailing patients based on their basic demography, diet, health history and other related data. The deep learning classifiers considered in this research include LSTM, MLP, GRU, RNN, Naive Bayes, and LR. The results of the performed experiments show that LSTM and GRU outperform every other classifier with reference to their recall, precision, and *F*-1 measures for both the allowed and not allowed classes. The performance of the LSTM classifier for the allowed class is measured at 100, 99, and 99% for precision, F1-measure, and recall, respectively. Meanwhile, the disallowed class performance of the LSTM classifier measures 90, 80, and 45% for recall, precision, and F1-measure, respectively. In the future, we shall incorporate the concept of multidimensionality in the secured diet recommender systems, where patients with special needs will maintain privacy after the whole process.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This paper was supported by the key research and development plan (social development) projects BE2016630 and BE2017628 of Jiangsu province, the scientific research project Z201603 of Wuxi health and family planning commission.

REFERENCES

1. Bourquin J, Schmidli H, Van Hoogevest P, Leuenberger H. Basic concepts of artificial neural networks (ANN) modeling in the application to pharmaceutical development. *Pharm DevTechnol.* (1997) 2:95–109.
2. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J Pharm Biomed Anal.* (2000) 22:717–27. doi: 10.1016/S0731-7085(99)00272-1
3. Puri M, Pathak Y, Sutariya VK, Tipparaju S, Moreno W. Introduction to artificial neural network (ANN) as a predictive tool for drug design discovery. In: *Delivery and Disposition: Basic Concepts and Modeling.* New York, NY: Academic Press (2016). p. 3–13.
4. Iwendi C, Suresh P, Revathi M, Srinivasan K, Chang C-Y. An efficient and unique TF/IDF algorithmic model-based data analysis for handling applications with big data streaming. *Electronics.* (2019) 8:1331. doi: 10.3390/electronics8111331
5. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access.* (2020) 8:28462–74. doi: 10.1109/ACCESS.2020.2968537
6. Srivastava G, Dhar S, Dwivedi AD, Crichigno J. Blockchain education. In: *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE).* (2019). p. 1–5.
7. Kim J-H, Lee J-H, Park J-S, Lee Y-H, Rim K-W. Design of diet recommendation system for healthcare service based on user information. In: *Proceedings of the Fourth International Conference on Computer Sciences and Convergence Information Technology.* Seoul: IEEE (2009). p. 516–8.
8. Banerjee A, Nigar N. Nourishment recommendation framework for children using machine learning and matching algorithm. In: *Proceedings of the International Conference on Computer Communication and Informatics (ICCCI).* Coimbatore: IEEE (2019). p. 1–6.
9. Miranda R, Ferreira D, Abelha A, Machado J. Intelligent nutrition in healthcare and continuous care. In: *Proceedings of the International Conference in Engineering Applications (ICEA).* Sao Miguel: IEEE (2019). p. 1–6.
10. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access.* (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
11. Sookrah R, Dhowtal JD, Nagowah SD. A DASH diet recommendation system for hypertensive patients using machine learning. In: *Proceedings of the 7th*

International Conference on Information and Communication Technology (ICoICT). (2019). p. 1–6.
12. Madumidha S, Ranjani PS, Varsinee SS, Sundari P. Transparency and traceability: In food supply chain system using blockchain technology with Internet of Things. In: *Proceedings of the 3rd International Conference on Trends in Electronics and Informatics (ICOEI).* (2019). p. 983–7.
13. Alian S, Li J, Pandey V. A personalized recommendation system to support diabetes self-management for American Indians. *IEEE Access.* (2018) 6:73041–51. doi: 10.1109/ACCESS.2018.2882138
14. Sahoo AK, Pradhan C, Barik RK, Dubey H. DeepReco: deep learning-based health recommender system using collaborative filtering. *Computation.* (2019) 7:25. doi: 10.3390/computation7020025
15. Jiang L, Yang CC. User recommendation in healthcare social media by assessing user similarity in heterogeneous network. *Artif Intell Med.* (2017) 81:63–77. doi: 10.1016/j.artmed.2017.03.002
16. Portugal I, Alencar P, Cowan D. The use of machine learning algorithms in recommender systems: a systematic review. *Expert Syst Appl.* (2018) 97:205–27. doi: 10.1016/j.eswa.2017.12.020
17. Panda SK, Blome A, Wisniewski L, Meyer A. IoT retrofitting approach for the food industry. In: *Proceedings of the 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA).* Zaragoza: IEEE (2019). p. 1639–42.
18. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol Forecasting Social Change.* (2018) 126:3–13. doi: 10.1016/j.techfore.2015.12.019
19. Yang CC, Jiang L. Enriching user experience in online health communities through thread recommendations and heterogeneous information network mining. *IEEE Trans Comput Soc Syst.* (2018) 5:1049–60. doi: 10.1109/TCSS.2018.2879044
20. Kaur H, Kumar N, Batra S. An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system. *Future Gener Comput Syst.* (2018) 86:297–307. doi: 10.1016/j.future.2018.03.017
21. Mu R. A survey of recommender systems based on deep learning. *IEEE Access.* (2018) 6:69009–22. doi: 10.1109/ACCESS.2018.2880197
22. Priyadarshini R, Barik R, Dubey H. DeepFog: Fog computing-based deep neural architecture for prediction of stress types diabetes and hypertension attacks. *Computation.* (2018) 6:62. doi: 10.3390/computation6040062
23. Agapito G, Calabrese B, Guzzi PH, Cannataro M, Simeoni M, Care I, et al. DIETOS: a recommender system for adaptive diet monitoring and personalized food suggestion. In: *Proceedings of the IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).* New York, NY: IEEE (2016). p. 1-8.
24. Baldominos Gomez, Rada F, Saez Y. DataCare: big data analytics solution for intelligent healthcare management. *Int J Interact Multimedia Artif Intell.* (2017) 4:13. doi: 10.9781/ijimai.2017.03.002
25. Leipold N, Lurz M, Bohm M. Nutrilize a personalized nutrition recommender system: an enable study. *HealthRecSys.* (2018) 3:4–10.
26. Jenzer H, Busser S, Scheidegger-Balmer F. Nutrition and dietetic product shortages are a neglected issue in alerting systems and in registries. *J Clin Nutr.* (2016) 2:15. doi: 10.4172/2472-1921.100022
27. Yera Toledo R, Alzahrani AA, Martinez L. A food recommender system considering nutritional information and user preferences. *IEEE Access.* (2019) 7:96695—711. doi: 10.1109/ACCESS.2019.2929413
28. Maiyaporn P, Phathrajarin P, Suphakant P. Food recommendation system using clustering analysis for diabetic patients. In: *Proceedings of the International Conference on Information Science and Applications,* Vol. 6. Seoul: IEEE (2010). p. 5–14.
29. Iwendi C, Anajemba JH, Biamba C, Ngabo D. Security of things intrusion detection system for smart healthcare. *Electronics.* (2021) 10:1375. doi: 10.3390/electronics10121375
30. Norouzi S, Ghalibaf AK, Sistani S. A mobile application for managing diabetic patients' nutrition: a food recommender system. *Arch Iranian Med.* (2018) 21:466–72.
31. Rehman F, Khalid O, Haq N, Khan AR, Sajjad KB. Diet-right: a smart food recommendation system. *KSII Trans Internet Inf Syst.* (2017) 11:2910–25.
32. Anajemba JH, Iwendi C, Mittal M, Yue T. Improved advance encryption standard with a privacy database structure for IoT nodes. In: *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT).* Gwalior (2020). p. 201–6.
33. Schäfer H, Elahi M, Elsweiler D, Groh G, Harvey M, Ludwig B, et al. User nutrition modelling and recommendation: balancing simplicity and complexity. In: *Proceedings of the Adjunct Publication of the 25th Conference on USER MODELING, Adaptation and Personalization (UMAP).* (2017). p. 9–12.
34. Kale A, Auti N. Automated menu planning algorithm for children: food recommendation by dietary management system using ID3 for Indian food database. *Procedia Comput Sci.* (2015) 50:197–202. doi: 10.1016/j.procs.2015.04.070
35. Anajemba JH, Yue T, Iwendi C, Alenezi M, Mittal M. Optimal cooperative offloading scheme for energy efficient multi-access edge computation. *IEEE Access.* (2020) 8:53931–41. doi: 10.1109/ACCESS.2020.2980196
36. Yang L, Hsieh C-K, Yang H, Pollak JP, Dell N, Belongie S, et al. Yum-Me: a personalized nutrient-based meal recommender system. *ACM Trans Inf Syst.* (2017) 36:1–31. doi: 10.1145/3072614
37. Anajemba JH, Tang Y, Iwendi C, Ohwoekevwo A, Srivastava G, Jo O. Realizing efficient security and privacy in IoT networks. *Sensors.* (2020) 20:2609. doi: 10.3390/s20092609
38. Saini S, Dubey SK. Recommendation of diet to jaundice patient on the basis of nutrients using AHP and fuzzy AHP technique. *Int J Intell Eng Syst.* (2017) 10:91–9. doi: 10.22266/ijies2017.0831.10
39. Chensi C, Feng L, Hai T, Deshou S, Wenjie S, Weizhong L, et al. Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinf.* (2018) 16:17–32. doi: 10.1016/j.gpb.2017.07.003
40. Trang Tran TN, Atas M, Felfernig A, Stettinger M. An overview of recommender systems in the healthy food domain. *J Intell Inf Syst.* (2018) 50:501–26. doi: 10.1007/s10844-017-0469-0
41. Kumar P, Kumar R, Srivastava G, Gupta GP, Tripathi R, Gadekallu TR, et al. PPSF: a privacy-preserving and secure framework using blockchain-based machine-learning for IoT-driven smart cities. In: *IEEE Transactions on Network Science and Engineering.*
42. Prabadevi B, Deepa N, Pham QV, Nguyen DC, Reddy T, Pathirana PN, et al. Toward blockchain for edge-of-things: a new paradigm, opportunities, and future directions. In: *IEEE Internet of Things Magazine 4.2.* (2021). p. 102–8.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mantey, Zhou, Anajemba, Okpalaoguchi and Chiadika. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fpubh.2021.744100

Recurrent Neural Network and Reinforcement Learning Model for COVID-19 Prediction

R. Lakshmana Kumar[1], Firoz Khan[2], Sadia Din[3], Shahab S. Band[4], Amir Mosavi[5,6]* and Ebuka Ibeke[7]*

[1] Department of Computer Applications, Hindusthan College of Engineering and Technology, Coimbatore, India, [2] Dubai Men's College, Higher Colleges of Technology, Dubai, United Arab Emirates, [3] Department of Information and Communication Engineering, Yeung University, Gyeongsan, South Korea, [4] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliu, Taiwan, [5] Faculty of Civil Engineering, Technische Universität Dresden, Dresden, Germany, [6] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [7] School of Creative and Cultural Business, Robert Gordon University, Aberdeen, United Kingdom

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Joseph Henry Arinze Anajemba,
Hohai University, China
Mohammad Al Rawajbeh,
Al-Zaytoonah University of Jordan, Jordan
Agbotiname Lucky Imoize,
University of Lagos, Nigeria
Sweta Bhattacharya,
VIT University, India

***Correspondence:**
Shahab S. Band
shamshirbands@yuntech.edu.tw
Amir Mosavi
amir.mosavi@mailbox.tu-dresden.de

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 19 July 2021
Accepted: 02 September 2021
Published: 04 October 2021

Citation:
Kumar RL, Khan F, Din S, Band SS, Mosavi A and Ibeke E (2021) Recurrent Neural Network and Reinforcement Learning Model for COVID-19 Prediction. Front. Public Health 9:744100. doi: 10.3389/fpubh.2021.744100

Detection and prediction of the novel Coronavirus present new challenges for the medical research community due to its widespread across the globe. Methods driven by Artificial Intelligence can help predict specific parameters, hazards, and outcomes of such a pandemic. Recently, deep learning-based approaches have proven a novel opportunity to determine various difficulties in prediction. In this work, two learning algorithms, namely deep learning and reinforcement learning, were developed to forecast COVID-19. This article constructs a model using Recurrent Neural Networks (RNN), particularly the Modified Long Short-Term Memory (MLSTM) model, to forecast the count of newly affected individuals, losses, and cures in the following few days. This study also suggests deep learning reinforcement to optimize COVID-19's predictive outcome based on symptoms. Real-world data was utilized to analyze the success of the suggested system. The findings show that the established approach promises prognosticating outcomes concerning the current COVID-19 pandemic and outperformed the Long Short-Term Memory (LSTM) model and the Machine Learning model, Logistic Regresion (LR) in terms of error rate.

Keywords: COVID-19, deep learning, LSTM, RNN, prediction reinforcement learning

INTRODUCTION

With the spread of the unfamiliar Coronavirus (COVID-19), which was first discovered in Wuhan city in China in 2019, societies worldwide continue to face very distressing times. On March 11, 2020, the World Health Organization (WHO) flagged the COVID-19 as a pandemic, exceeding 118,000 cases in over 110 countries. The epidemic has quickly spread through many countries, including Italy, Spain, France, the United States, and India, wreaking havoc on healthcare systems (1). Modeling and predicting the expanse of verified and recovered COVID-19 cases accurately is critical for understanding and helping decision-makers to slow down or arrest its progression. Since the COVID-19 pandemic has shifted into a global pandemic, there is a necessity for real-time epidemiological data examinations to provide the population with a strong course of action to combat the infection. Following the novel COVID-19, the world has been restlessly battling its cause (2). As of August 27, 2020, there were 24,631,906 confirmed cases worldwide, of which

TABLE 1 | Top 5 affected countries COVID-19 data.

Country	Cases	Deaths	Recovered	Fatality (%)
USA	28,765,423	511,133	18,973,190	1.8
India	11,005,850	156,418	10,699,410	1.4
Brazil	10,168,174	246,560	9,095,483	2.4
Russia	4,164,726	83,293	3,713,445	2.0
UK	4,115,509	120,580	2,494,218	2.96

TABLE 2 | Top 5 affected countries COVID-19 data.

Name	COVID-19	MERS	SARS
Year	2019	2012	2003
Confirmed	110,927,514	2,494	8,096
Deaths	2,454,587	858	774
Fatality %	2.21	34.1	9.6
Countries	220	27	26

17,089,939 recovered, and 841,310 ended in death (3). **Table 1** shows the topmost countries affected. The COVID-19 relate itself to the species as that of SARS-CoV and MERS-CoV.

Table 2 shows the comparison, where the symptoms initially appear as a common cold then progress to those of respiratory diseases that cause breathing problems, tiredness, fever, and dry cough. Once a large-scale break out of a contagious disease occurs and a significant public health emergency ensues, researchers use outbreak models to evaluate and forecast the disease's development pattern and determine direct measures to prevent and restraint based on the effects of the analysis.

The most frequently used conventional pandemic schemes are susceptible—infected—recovered (SIR), and susceptible—exposed—infected—recovered (SEIR) models (4), where "S," "E," "I," and "R" signify every number of susceptive persons, the magnitude of individuals during the incubation phase, the magnitude of contagious persons and the number of individuals improved, respectively. These models are trained to forecast multiple diseases, such as Ebola and SARS, due to their robust predictive abilities of the linked indications. With the emergence and spread of COVID-19, the significant research challenge is arresting the growth patterns of the spread of this disease which has been observed in several science fields throughout the globe. Thereby, different approaches (5) to modeling, estimating, and forecasting are implemented to understand and control this pandemic. Traditional disease models measure the rate of infection based on the complex variation in the number of contaminations and then determine the disease's spread and evolution pattern. Yet, those approaches assume that all individuals with Coronavirus hold an equal chance of infection, and hence, their predictive results can only suggest general patterns and are restricted.

Artificial Intelligence (AI) is lately being applied toward stimulating biomedical study and toward numerous fields such as image identification, object categorization, image segmentation, and deep learning approaches (6). For example, individuals affected with COVID-19 will possibly have pneumonia since the infection reaches the lungs. Many deep learning investigations identify the condition using X-ray images of the chest (7). Three different deep learning models (8) have been employed in the past to distribute X-ray images of pneumonia, and those are the fine-tuned model, the non-fine-tuned and the scratch-trained model. On the other hand, most prediction models use X-ray and CT images (9) based on the deep learning method, which requires more time to extract the features and train the model.

Famous classical mathematical differential equations and population prediction models have limitations on predicting the population in the time-series and significant estimation errors. Analytical methodologies, for instance, Auto Regressive Moving Average (ARIMA), Moving average (MA), and Auto-Regressive (AR) methods, are primarily formulated on the premises. Still, they have difficulties in predicting live circulation rates. A vast variety of demographic and computative models (10) were developed for modeling COVID-19's rampant transmission dynamics. However, in multiple situations, these approaches don't adhere to the provided information, and the accuracy of the forecast is usually low. Therefore, this work investigates the modified LSTM approach to forecasting the likely COVID-19 cases and deaths. It also describes deep reinforcement learning for optimizing the prediction results based on symptoms. Experiments using real data and various metrics reveal the improved performance of the work. The specific contributions of this paper include:

- Deep and reinforcement learning to predict COVID-19,
- LSTM model modified with new activation function for efficient prediction,
- Deep reinforcement learning applied to optimize results based on COVID-19 symptoms

The remainder of the paper is ordered as follows. Section related work reviews the related work. Preliminary information regarding the used approaches and the problem statement is given in section methodological preliminaries, while section optimized prediction of covid-19 describes the proposed method. The experiential details, evaluation criteria, and performance comparison are given in section optimized prediction of covid-19, along with an analysis of the conclusions attained. Lastly, section results and discussions gives some concluding observations.

RELATED WORK

COVID-19 Prediction and Forecasting

Various prediction techniques that are regularly used to tackle forecasting problems include Machine Learning (ML) models (4), which can be employed toward determining the number of potential COVID-19 infected patients. Rustam et al. (4) used four simple statistical models: Linear Regression, Least Absolute Shrinkage and Selection Operator, Support Vector Machine, and Exponential Smoothing to forecast threatening COVID-19 factors. Petropoulos and Makridakis (11) presented an analytical procedure to forecast the continuation of COVID-19. The work presents a timeline of a live forecasting activity with

significant possible planning and decision-making consequences and offers realistic forecasts for confirmed COVID-19 cases. Malavika et al. (12) adopted a logistic growth curve model for short-term prediction of COVID-19, and SIR models was employed in identifying the highest possible live individuals and peak seasons. In addition, the Time Disrupted Regression model is used to estimate the influence of lockdown and other important proposals.

Pal et al. (13), combined medical data with the trend and local weather data to forecast each country's level of risk. Specifically, a shallow LSTM neural network is employed in solving difficulties in limited datasets, and a country's risk level (high, medium, and recovery) is categorized using the Fuzzy rule. Hu et al. (14) used Coronavirus-specific dataset to fine-tune the pre-trained multi-task deep model. The re-trained prototype was then utilized to decide possible commercial medications upon targeted proteins of SARS-CoV-2. Finally, Salgotra et al. (9) developed Genetic programming (GP) prediction models for confirmed individuals and death cases across three of the most affected states, namely Maharashtra, Gujarat, Delhi, and India. The predictive models are expressed utilizing the specific formula, and predictive powerless variables were studied.

Velásquez and Lara (1) analyzed historical and expected COVID-19 death infections based upon the Reduced-Space Gaussian Process Regression correlated with disordered Dynamical Systems. COVID-19 forecasted with Gaussian models mean-field models can be meaningfully applied to obtain a quantitative summary of virus spread with contamination, death, and recovery rates. Jia et al. (10) adopted Logistic, Bertalanffy, and Gompertz models to prove the validity of the current statistical models by fitting and analyzing the SARS epidemic patterns. The findings were then used to fit and evaluate the COVID-19 scenarios. The forecasted outcomes of the three different mathematical models varied for different parameters and in different regions. Kavadi et al., (15) proposed partial derivative regression and a non-linear machine learning system toward the global pandemic verification of COVID-19. Dehesh et al., (16) considered the best predictive models for regularly reported individuals in nations with a huge magnitude of verified cases and then made predictions based on those models to better prepare healthcare systems. For predicting the pattern of reported events, the Auto-Regressive Integrated Moving Average model was used. Ngabo et al. (17) proposed an artificial intelligence (AI) algorithm that predicts the survival rate of COVID-19 patients based on their immune system, exercise rate, and age quantiles.

Deep Learning for COVID-19

Arora et al. (18) used deep learning-based models to forecast the number of recorded positive cases of novel Coronavirus (COVID-19) for 32 Indian states and union territories. Recurrent Neural Network (RNN)-based LSTM variants, such as Deep LSTM, Convolutional LSTM and Bi-directional LSTM, were applied to the Indian dataset to forecast the number of positive cases. Huang et al. (19) suggest that Convolutional Neural Network (CNN) can accurately estimate and determine the number of verified cases. The emphasis was on various towns with the most reported cases in China, and a COVID-19 prognostication model was suggested based upon the CNN system of Deep Neural Network (DNN). Three deep learning models (20), namely DNN, LSTM, and CNN, were stacked in learning models for the ensemble to achieve the most reliable results. The meta-learners used these forecasted values of these models as inputs to produce the final prediction of outbreaks. Ramchandani et al. (21) proposed a deep learning model to forecast the range of increase in COVID-19 and offer an unusual approach for determining equidimensional multivariate time scale illustrations and multivariate spatial time scale results.

Yoo et al. (22) examined the usefulness of applying a deep learning-based decision-tree classifier to distinguish COVID-19 from CXR images. This classifier consists of three binary decision trees, each trained by a deep learning model based on the PyTorch system with a neural convolution network. The primary decision tree divides the CXR images as either regular or anomalous. The second tree recognizes the irregular images bearing symptoms of tuberculosis. The final tree identifies the signs of COVID-19. Ozturk et al. (23) introduced a different approach for automated COVID-19 detection employing raw X-ray images of the chest. The developed system offers honest diagnostics for binary (COVID-19 vs. No-Findings) and multi-class (COVID-19 vs. Pneumonia vs. No-Findings) classification. Panwar et al., (24) proposed a deep learning neural network-based approach nCOVnet, which employs an alternative rapid screening system to identify COVID-19 by analyzing patients' X-rays to check for visual markers present in COVID-19 patients' chest radiography images.

Hu et al. (25) proposed the weakly controlled deep learning strategy for recognizing and distinguishing COVID-19 contagion from computed tomography (CT) images. This approach reduces the manual labeling requirements for CT images, reliably diagnose infection, and distinguish COVID- 19 from non-COVID-19 cases. Deep learning-based research about CT in the chest has proven to be reliable and effective for determining COVID-19. Mohammed et al. (26) proposed ResNext+, which offers an end-to-end semi-supervised strategy to COVID-19 discovery, including data labels at volume level only, and can provide a slice level prediction. A deep, long-term bidirectional memory network with a mixed density network (DBM) model was established by Pathak et al. (27), namely the Memetic Adaptive Differential Evolution (MADE) algorithm which can fine-tune the hyperparameters for the DBM model.

Deep Reinforcement Learning

Reinforcement Learning (RL) is a machine-learning model, where agents learn efficient techniques from trial-and-error encounters with their surroundings that produce the single most massive, long-term reward. The Q-learning algorithm (28) is the most descriptive of the RL algorithms. Q-learning can learn an acceptable method without an environmental operating prototype by modifying an action-value algorithm called the Q function. When the state-action space is large and complex, deep neural networks can approximate the Q-equation, and the corresponding algorithm is called Deep Reinforcement Learning

(DRL) (29). This has promising application for rational decision-making in diverse fields, such as energy management, robotics, agriculture, healthcare, etc. This model successfully resolved a wide range of complicated decision-making assignments which were earlier outside the machine's limits.

Wang et al. (30) offered an adaptive design that relies on graph embedding in the training process during state representation and reinforcement learning. Depending on a couple of real-life datasets, the findings show that the scheme can beneficially decrease the infection's epidemiological replication rate. This approach can aid in the initial exposure of COVID-19, whereby RL may represent an effective method to combat the spread of an outbreak. Dell'Aversana (23) combines multi-layer Artificial Neural Networks with Reinforcement Learning architecture to allow software-defined factors to acquire environmentally optimized functioning. Iwendi et al., (31) utilized COVID-19 patients' geographical and demographic data to predict the severity of cases, recovery, and death. In (32) a semantic privacy framework that uses sensitive and semantically related terms to sanitize healthcare documents was proposed, and (33) uses deep learning to detect and sanitize social media comments. In (34), the authors discussed the concepts of an incentive approach for COVID-19 planning using Blockchain Technology. Deep learning and medical image processing for Coronavirus (COVID-19) pandemic were analyzed by the authors in (35) and the results are nicely presented.

METHODOLOGICAL PRELIMINARIES

This section describes the COVID-19 prediction problem and gives information about the technical background of DL and RL used within this work.

Problem Formulation

Time scale forecasting aims to utilize the input sequences witnessed earlier to forecast a fixed-length series of expected time scale values. In machine learning, a part of the input time-series sequence, i.e., delayed values, is replaced to assist the input functions. The number of leading time levels is recognized as the width/size of the frame. Provided with a single variable time-series:

$$TS\,(t) = \{s_1, s_2, s_3, \ldots, s_t\} \tag{1}$$

the intention is to forecast the future k values of the sequence, $\hat{y} = \hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots, \hat{y}_k \cong \left(s_{t+1}, s_{t+2}, s_{t+3}, \ldots, s_{t+k}\right)$ utilizing the values of former conclusions.

Long Short-Term Memory Network

RNN is one of the deep learning techniques which automatically selects appropriate characteristics from the practice specimens and later supplies activation from the previous time step as information for the current time step and network's self-connections. RNN is proper for data processing and has outstanding potential in time-series forecast by saving extensive historical data in its inner state. Still, it has the limitation of disappearing and gradient-exploding difficulties, which leads to an extended practice period or practice that does not work. In 1997, Hochreiter and Schmidhuber (36) devised a long short-term memory structure to determine long-term dependence on the multiplicative passages that coordinate information and memory cells movement in the recurrent hidden layer.

Figure 1 exhibits an LSTM memory cell's primary arrangement with two distinct components: (C_t) and the short-term state component (h_t).

The structure of LSTM consists of the following gates: input, forget, control, and output gate. The input gate determines which information can be passed on to the cell and is defined as:

$$i_t = \sigma\, W_i \times h_{t-1} x_t + b_i \tag{2}$$

The information to be ignored from the previous memory input is determined by the forget gate and is defined as:

$$f_t = \sigma\, W_f \times h_{t-1} x_t + b_f \tag{3}$$

The cell update is controlled by the control gate, based on the following equations:

$$\tilde{C}_t = \tanh W_c \widetilde{\times} h_{t-1} X_t + b_f \tag{4}$$

$$C_t = F_t^* \tilde{C}_{t-1} + i_t^* \tilde{C}_t \tag{5}$$

The hidden layer (h_{t-1}) is updated by the output layer, which is also responsible for updating the output gate as is given by:

$$O_t = \sigma\, W_o \times h_{t-1} x_t + b_o \tag{6}$$

In the above equation, W and b represent the weight matrix and bias vector, respectively; tanh is used to scale the values in the range −1 to 1, and σ denotes a standard logistic sigmoid function. The variables *i, f, o,* and *c* are the input gate, forget gate, output gate, and cell activation vector.

Reinforcement Learning

Reinforcement learning (RL) is an artificial intelligence model with a progressive programming guide instructing algorithms to apply an award and penalty strategy. A Markov choice system is designated for the RL study process, which endorses the formalism of reinforcement learning difficulties. The RL algorithm is an agent that receives through communicating and associating with society. The agent will obtain incentives for the appropriate steps and penalties for the inaccurate performances. Without individual intervention, the agent learns by itself via improving his incentives and reducing his punishments. The intercommunication process between the agent and environment of RL is shown in **Figure 2**.

An agent available in a state (S) executes an action (A). The agent collects a reward R(S, A) for acting and shifts into a different state. A policy (π) means any mapping function of the states and actions that decide an agent's action in every state. The central goal in an agent's existence is to obtain an optimal

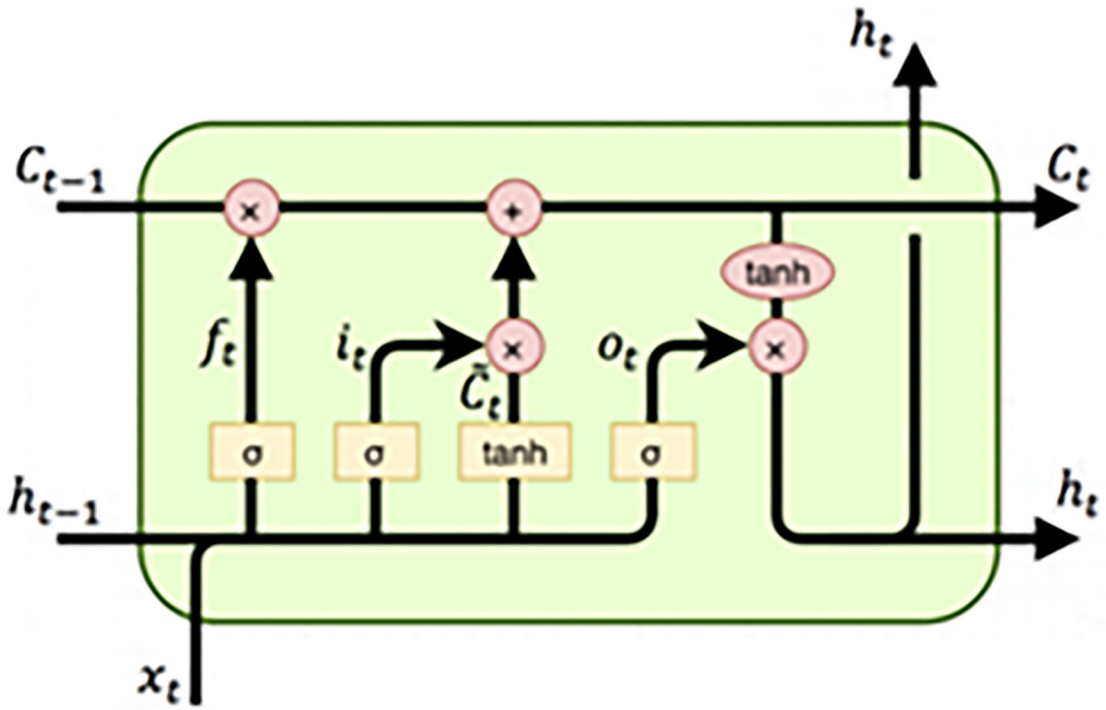


FIGURE 1 | The structure of the LSTM.

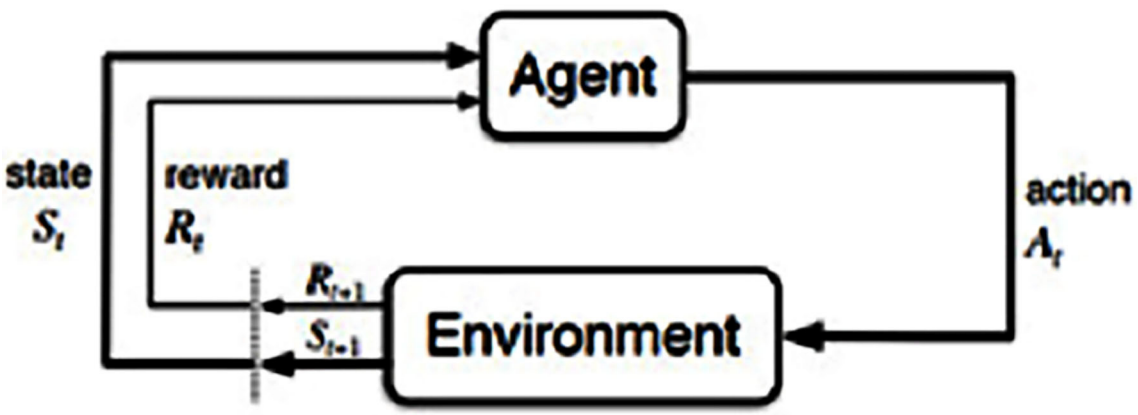


FIGURE 2 | Interaction process of reinforcement learning.

strategy π^* for which the overall reduced incentive is expanded. The optimal system π^* is defined in equation (7):

$$\pi^*(S) = argmax_{a\epsilon A}\gamma \sum_{s'\epsilon S} P_{sa}\left(s', a\right) V^*(s', a) \tag{7}$$

The best assessment function is obtained from the most dependable optimal plan; it is defined through some benefits an agent receives from all the other states. The aforementioned function with optimum power is expressed in equation (8):

$$V^*(s, a) = R(s, a) + max_{a\epsilon A}\gamma \sum_{s'\epsilon S} P_{sa}\left(s', a\right) V^*\left(s', a\right) \tag{8}$$

Thus, a reinforcement learning agent learns through experiences with the environment. Using complex programming functions, agents optimize their compensations through estimating individual fittest optimal strategy and power function for the bellman.

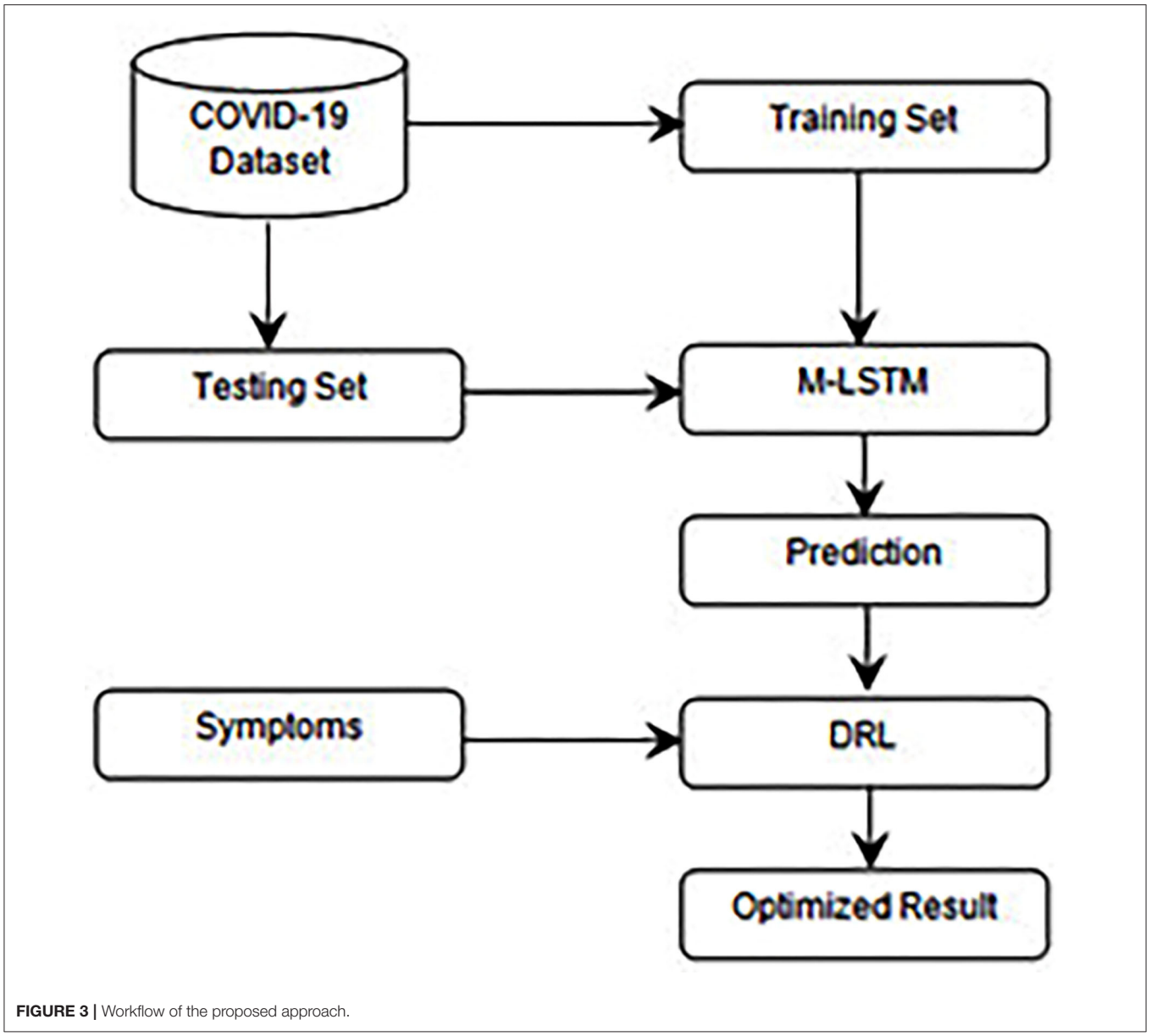


FIGURE 3 | Workflow of the proposed approach.

OPTIMIZED PREDICTION OF COVID-19

The presented research investigates predictions for the novel COVID-19 using the discussed reinforcement learning. COVID-19 has proved to be a serious threat to people of all ages all over the world. It has resulted in tens of thousands of deaths so far, with the mortality rate continuing to rise regularly. This research sought to assess future estimates of the death rate, the number of infected individuals reported every day and the number of healing cases in the following 15 days to add to the pandemic's tracking. The forecasting was done using a Deep Learning framework that was designed specifically for this study.

AI is a growing platform with many different intelligent applications, such as Machine Learning (ML) and Deep Learning (DL). ML indicates the capability to acquire and deduce significant patterns of these data; furthermore, ML-based algorithms and practices' accomplishment depends heavily on the individual functions. Meanwhile, by learning from simple representation, DL can solve complex systems. DL possesses a pair of critical features (14): (1) the capacity to acquire the correct phrases and (2) the ability to help the machine discover data by sequentially using several layers to understand more meaningful representations.

This paper proposes an optimized prognostication of COVID-19 employing deep reinforcement learning (DRL). Primarily, Modified-LSTM was utilized to forecast the figures of verified cases and death cases. Next, the forecasted outcomes were optimized by employing

TABLE 3 | Daily and weekly sample data for COVID-19.

Date	Confirmed	Deaths	Weekly confirmed	Weekly deaths
30–01–20	1	0	1	0
31–01–20	0	0	1	0
01–02–20	0	0	1	0
02–02–20	1	0	2	0
…	…	…	…	…
14–08–20	64,553	1,007	434,116	6,455
15–08–20	65,002	996	437,581	6,518
16–08–20	63,490	944	436,672	6,601

DRL based on the symptoms. **Figure 3** shows the proposed workflow.

LSTM is an RNN deformation structure, where the memory cell is added to the hidden layer to restrain the memory data of the timeline data. Data is transmitted through several controllable gates between different cells of the hidden layer, thus allowing control of the memory and ignoring the amount of the former and prevailing data. Two LSTM gates are designed to control the status of the memory cell. One is the forget gate, which shows how much "memory" can be preserved from the last moment of the cell; the second is the input gate, which specifies how much information about the current time can be kept to the cell status and regulates the fusion ratio of "old" data and "present" incentive. Lastly, LSTM's output gate is designed to control the extent of cell status information outputs. The structure of LSTM can be found in **Figure 1**.

Herein, two activation functions of LSTM are used: linear and non-linear. The traditional LSTM uses the non-linear tanh function. For best results, select the best activation function.

$$f(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \tag{9}$$

Deep learning stimulates the improvement in RL, whereby DL algorithms within RL describe the deep reinforcement learning (DRL) area. Deep learning allows RL to scale up the earlier unmanageable issues, i.e., settings by a high-dimensional state, areas for the interruption, and decision-making. Deep reinforcement training employs a deep neural network to approximate every reinforcement learning function, including value function, Q function, transformation system, and reward function. Q-Learning is an RL system that decides which action an agent should take, depending on an action-value role. This determines the significance of remaining in a particular state and completing a specific activity at that state depending on an action-value role.

It is one of the most meaningful progress in reinforcement learning by developing an algorithm to limit off-policy temporary deviation. Q-Learning measures a state-action value function for a target system that decides the highest value in picking the

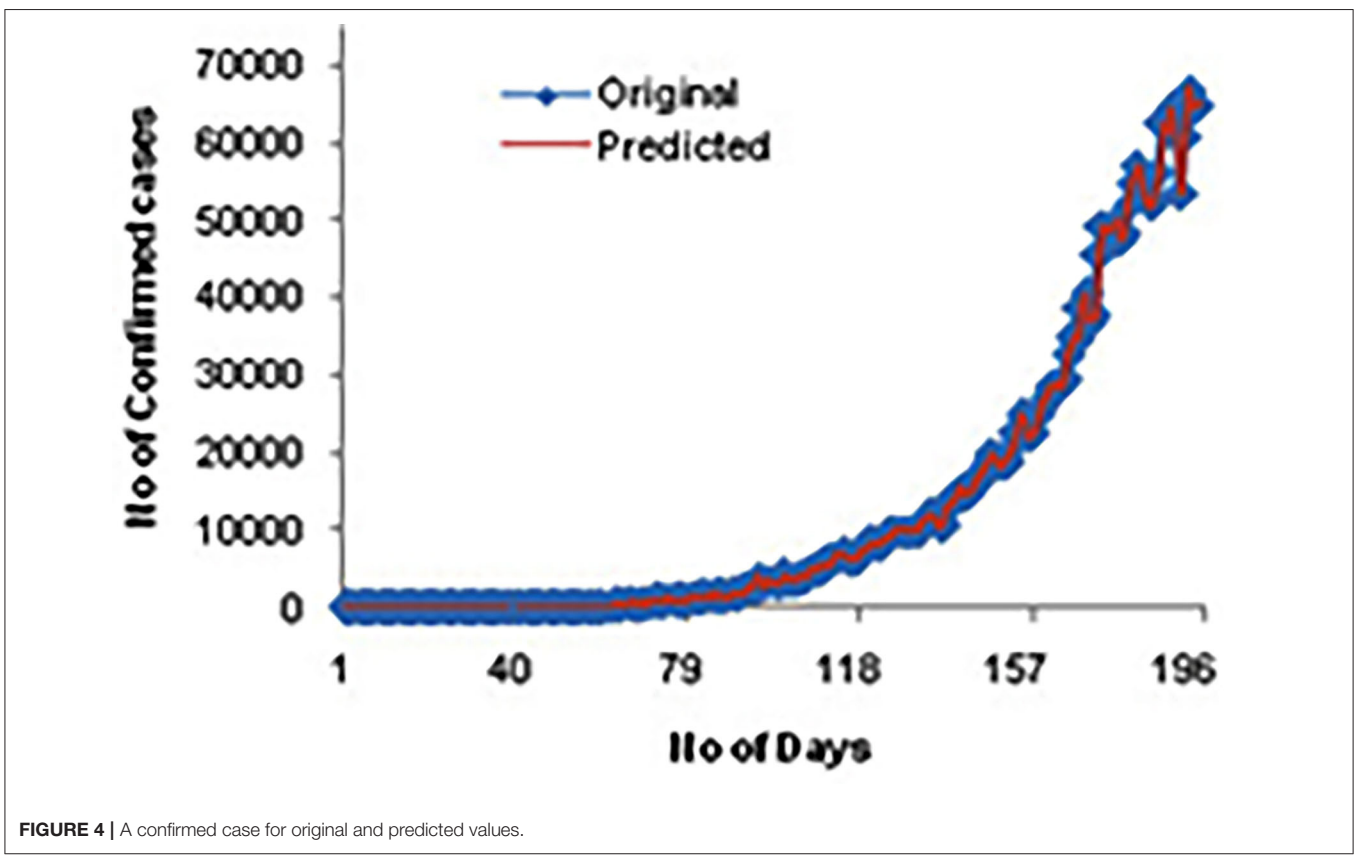


FIGURE 4 | A confirmed case for original and predicted values.

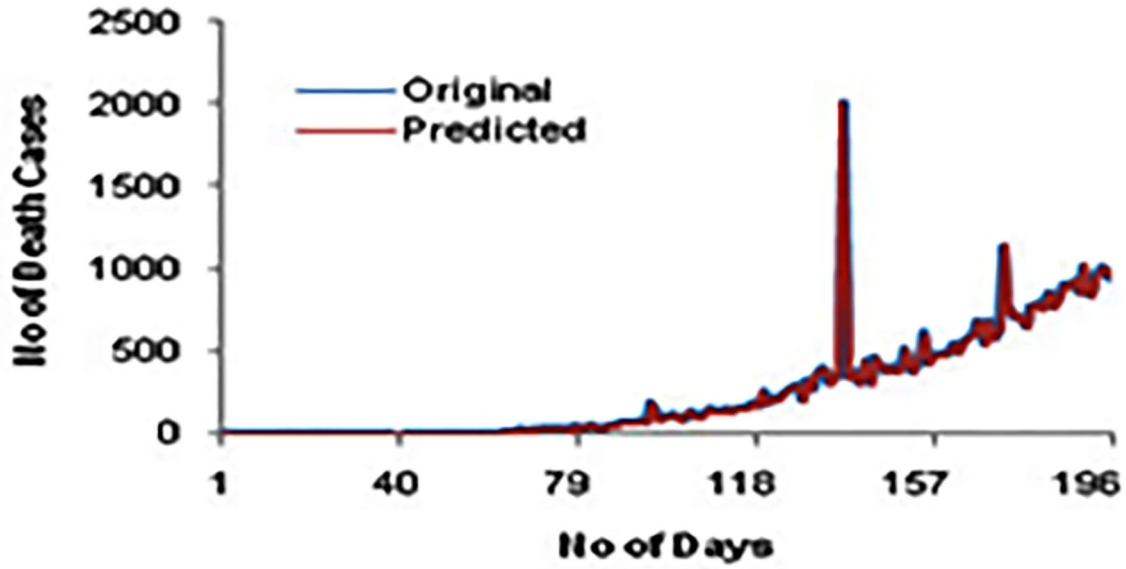


FIGURE 5 | Predicted values for death cases.

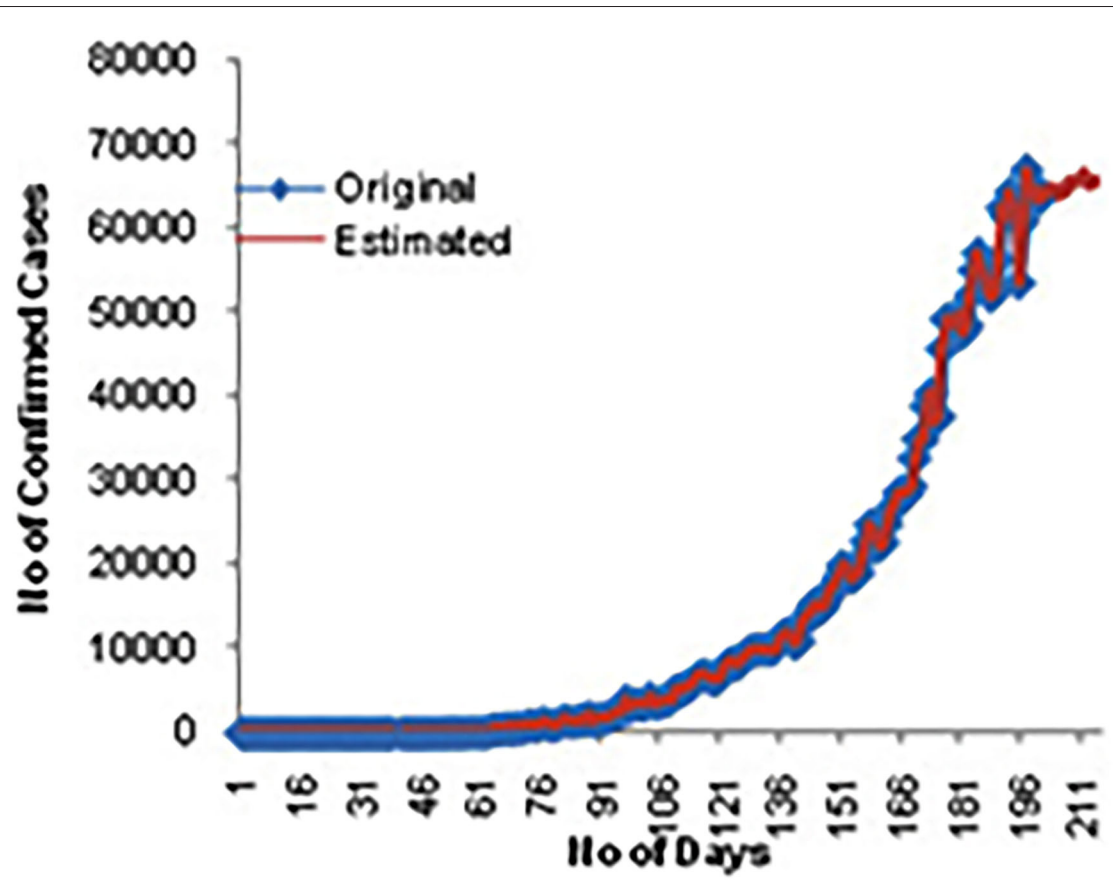


FIGURE 6 | Estimated confirmed cases for 15 days.

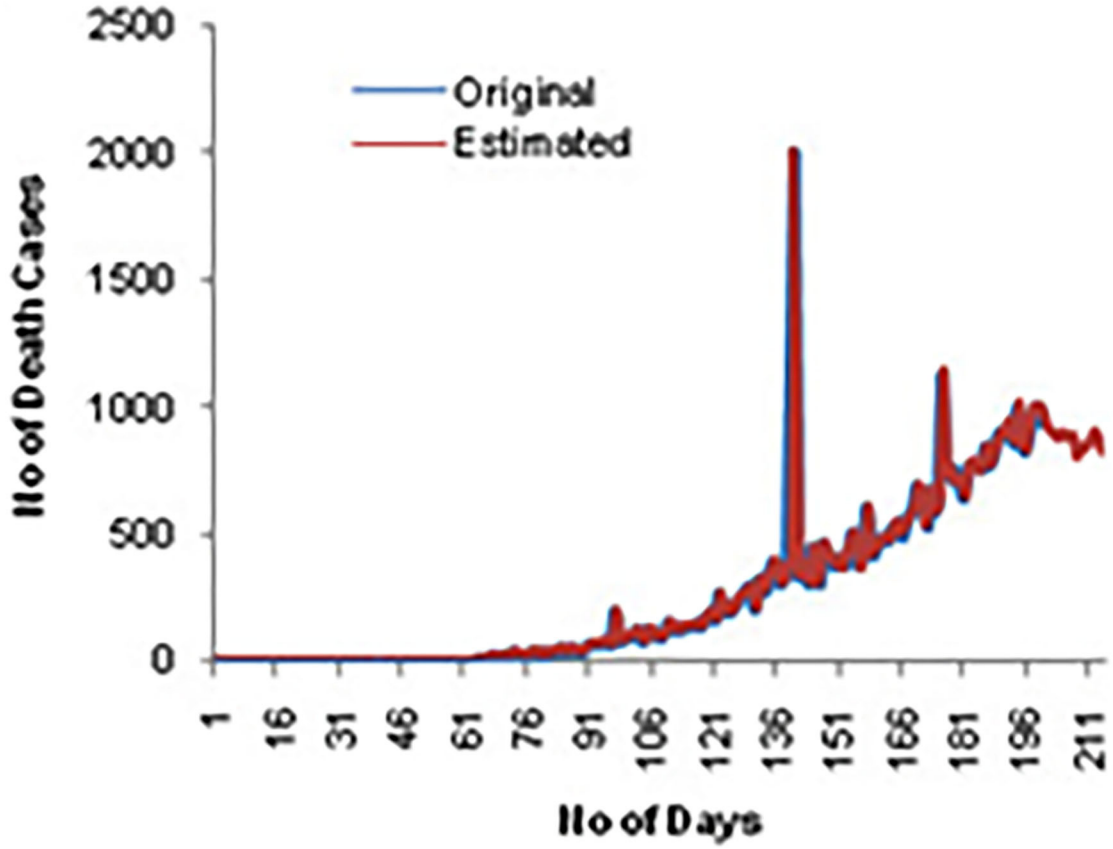


FIGURE 7 | Death case estimated for next 15 days.

action. Function Q accepts the information as a current state (S) and action (A) and returns an estimated reward for that action in that state. Thus, q functions provide the arbitrary fixed values in the beginning before investigating the situation.

In this work, the following symptoms were considered to predict COVID-19 cases: fever, tiredness, dry cough, difficulty breathing, sore throat, pains, nasal congestion, runny nose, and diarrhea. Groups of symptoms are considered states, and action is taken based on the states. Here, the action is taken based on the increments in the confirmed and death cases. A reward is received when the foresight of verified cases and death cases are accurate. The action-value function defined as:

$$q(s, a; \theta) \approx Q^{*}(s, a) \tag{10}$$

where q (s, a) represents the neural network approximation and θ is a reference variable representing the network's edge weights. The input to the neural network is a state, and the outputs for unconnected activities $Q=\{q(s,a,\theta)|a\epsilon A\}s$ are approximate q values.

The system is trained by depreciating forecast faults of $q(s, a; \theta)$. The DRL agent taking action at time t is $a_t = \arg\max_a q(s_t, a; \theta)$, where $q(s_t, a; \theta)$ for various activities are supplied through the outcomes of the network. For example, assume that the resulting compensation is r_{t+1} and the state move to s_{t+1}, then various actions are provided by the outputs of QNN. Consider that the resulting reward is r_{t+1} and the state moves to s_{t+1}, then $(s_t,a_t,r_{t+1},s_{t+1})$ establishes an "experience sample" that might be used to train the network. For training, the prediction error of network for the particular experience sample $(s_t,a_t,r_{t+1},s_{t+1})$ is defined as:

$$L_{s_t}, a_t, r_{t+1}, s_{t+1}(\theta) = (y_{r_{t+1}}, s_{t+1} - q(s_t, a_t; \theta))^2 \tag{11}$$

Where $\theta = weights$; and $\left(y_{r_{t+1}}, s_{t+1}\right) = targetoutput$, which is defined as:

$$\left(y_{r_{t+1}}, s_{t+1}\right) = r_{t+1} + \gamma \max_{a'} q(s_{t+1}, a'; \theta) \tag{12}$$

RESULTS AND DISCUSSIONS

This segment displays the trial setup and evaluation manner of the suggested prognostication design. Data of COVID-19 cases in India from Kaggle were employed with the Java framework to authenticate and examine the recommended standard.

The dataset utilized in the report comprises summary tables of regular time scales, including the number of reported cases and

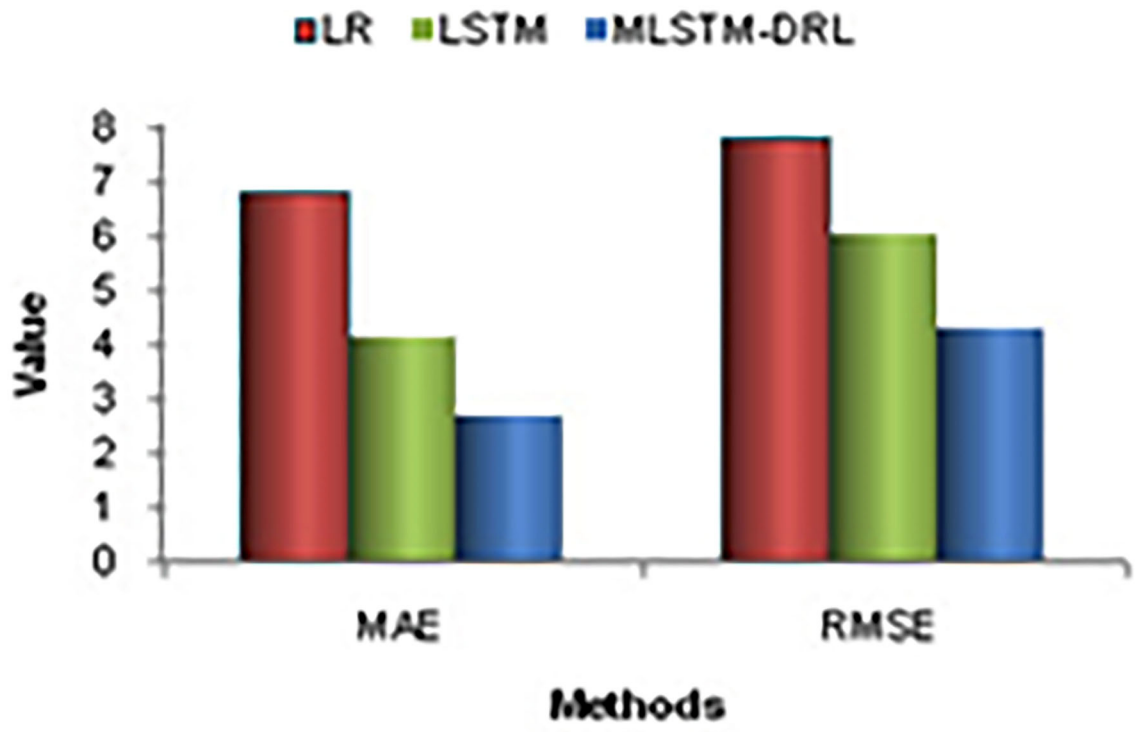


FIGURE 8 | Comparison of evaluation metrics.

deaths over the past number of days from which the pandemic began. Data from January 30, 2020 (when the first case of COVID-19 was registered in India) to August 16, 2020, were analyzed, with 75% data employed for practice and 25% for predictive and validation purposes. **Table 3** shows the sample data of confirmed and death cases daily and weekly.

Figure 4 shows the comparison of confirmed cases for the original and predicted values for 200 days. Here, we could see a high correlation between the actual and predicted cases.

Figure 5 shows the comparison of death cases for original and predicted values. It can be seen that the death rate slightly increased on specific days.

Figure 6 shows the estimated confirmed cases for the next 15 days. **Figure 7** displays the estimated death cases for the next 15 days.

The proposed algorithm can evaluate the performance of learning in terms of the following metrics: Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). **Figure 8** shows the comparison of evaluation metrics. Based on the outcomes, the recommended MLSTM-DRL has a lowest error rate matched to other systems. We also see the ML model, Logistic Regression (LR) obtained the highest error rate. Deep learning methods are utilized to develop a system for future prediction of the COVID-19 affected cases. The study performs predictions on confirmed and death cases. It is a troubling circumstance for the world day by day as death and reported cases are rising. The number of individuals in various countries affected by the COVID-19 pandemic is not well-known. This analysis attempts to estimate the number of individuals who will be affected over the next 15 days in terms of freshly authenticated cases and deaths.

Individuals being affected by COVID-19 get increased daily, and the death rate has many ups and downs. The correlation between original and predicted data of confirmed and death cases are 0.999. The future prediction results can assist governments in planning lockdown or other medical decisions.

CONCLUSION

COVID-19 is an ongoing pandemic that significantly endangers the health of people worldwide in a short period. A DL-based prediction method for forecasting the risk of COVID-19 has been proposed in this work. The framework analyses the actual day-to-day data dataset and uses deep learning algorithms to make predictions for upcoming days. This study determines the best activation function for M-LSTM; specifically, a deep reinforcement learning algorithm to optimize the prediction results. The proposed approach was compared with widely used existing algorithms like LR and LSTM. The finding of this work proves that the DL method can efficiently predict future cases of COVID-19. Overall, it can be concluded that the model's predictions are at per with the status of the virus; this may help understand and curb the spread of the virus.

Therefore, this study's forecast may be of great help in taking timely actions and making decisions to tackle the COVID-19 crisis. In the future, we advise using a semi-supervised hybrid

design to identify COVID-19 and social media platforms to prevent further spread. It is also planned to publish the predicted results as a dashboard through Google Data Cloud.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Open Access Funding by the Publication Fund of the TU Dresden.

REFERENCES

1. Velásquez RM, Lara JV. Forecast and evaluation of COVID-19 spreading in USA with reduced-space Gaussian process regression. *Chaos Solitons Fractals*. (2020) 136:109924. doi: 10.1016/j.chaos.2020.109924
2. Punn NS, Sonbhadra SK, Agarwal S. COVID-19 epidemic analysis using machine learning and deep learning algorithms. *MedRxiv*. (2020) 29:105340. doi: 10.1101/2020.04.08.20057679
3. *CoronaBoard: COVID-19 Dashboard*. Available online at: https://coronaboard.kr/ (accessed July 22, 2020).
4. Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, et al. COVID-19 future forecasting using supervised machine learning models. *IEEE Access*. (2020) 8:101489–99. doi: 10.1109/ACCESS.2020.2997311
5. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*. (2020) 12:165–74. doi: 10.21037/jtd.2020.02.64
6. Togaçar M, Ergen B, Cömert Z. Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders. *Med Hypotheses*. (2020) 135:109503. doi: 10.1016/j.mehy.2019.109503
7. Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ. Identifying pneumonia in chest X-rays: a deep learning approach. *Measurement*. (2019) 145:511–8. doi: 10.1016/j.measurement.2019.05.076
8. Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A. Comparison of deep learning approaches for multi-label chest X-ray classification. *Sci Rep*. (2019) 9:6381. doi: 10.1038/s41598-019-42294-8
9. Salgotra R, Gandomi M, Gandomi AH. Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos Solitons Fractals*. (2020) 138:109945. doi: 10.1016/j.chaos.2020.109945
10. Jia L, Li K, Jiang Y, Guo X. Prediction and analysis of coronavirus disease 2019. *arXiv [Preprint]* arXiv:2003.05447. (2020).
11. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. *PLos ONE*. (2020) 15:e0231236. doi: 10.1371/journal.pone.0231236
12. Malavika B, Marimuthu S, Joy M, Nadaraj A, Asirvatham ES, Jeyaseelan L. Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models. *Clin Epidemiol Glob Health*. (2021) 9:26–33. doi: 10.1016/j.cegh.2020.06.006
13. Pal R, Sekh AA, Kar S, Prasad DK. Neural network based country wise risk prediction of COVID-19. *Appl Sci*. (2020) 10:6448. doi: 10.3390/app10186448
14. Hu F, Jiang J, Yin P. Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model. *arXiv [Preprint]* arXiv:2003.00728. (2020).
15. Kavadi DP, Patan R, Ramachandran M, Gandomi AH. Partial derivative nonlinear global pandemic machine learning prediction of covid 19. *Chaos Solitons Fractals*. (2020) 139:110056. doi: 10.1016/j.chaos.2020.110056
16. Dehesh T, Mardani-Fard HA, Dehesh P. Forecasting of covid-19 confirmed cases in different countries with Arima models. *medrXiv*. (2020).
17. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng*. (2021).
18. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India. *Chaos Solitons Fractals*. (2020) 139:110017. doi: 10.1016/j.chaos.2020.110017
19. Huang CJ, Chen YH, Ma Y, Kuo PH. Multiple-input deep convolutional neural network model for covid-19 forecasting in china. *medRxiv*. (2020).
20. Yahia NB, Kandara MD, Saoud NBB. Deep ensemble learning method to forecast COVID-19 outbreak. (2020).
21. Ramchandani A, Fan C, Mostafavi A. Deepcovidnet: an interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions. *IEEE Access*. (2020) 8:159915–30. doi: 10.1109/ACCESS.2020.3019989
22. Yoo SH, Geng H, Chiu TL, Yu SK, Cho DC, Heo J, et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front Med*. (2020) 7:427. doi: 10.3389/fmed.2020.00427
23. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med*. (2020) 121:103792. doi: 10.1016/j.compbiomed.2020.103792
24. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos Solitons Fractals*. (2020) 138:109944. doi: 10.1016/j.chaos.2020.109944
25. Hu S, Gao Y, Niu Z, Jiang Y, Li L, Xiao X, et al. Weakly supervised deep learning for covid-19 infection detection and classification from ct images. *IEEE Access*. (2020) 8:118869–83. doi: 10.1109/ACCESS.2020.3005510
26. Mohammed AK, Wang C, Zhao M, Ullah M, Naseem R, Wang H, et al. Semi-supervised network for detection of COVID-19 in chest CT scans. *IEEE Access*. (2020) 8:155987–6000. doi: 10.1109/ACCESS.2020.3018498
27. Pathak Y, Shukla PK, Arya KV. Deep bidirectional classification model for COVID-19 disease infected patients. *IEEE/ACM Trans Comput Biol Bioinformat*. (2020). 18:1234–41. doi: 10.1109/TCBB.2020.3009859
28. Dai H, Khalil EB, Zhang Y, Dilkina B, Song L. Learning combinatorial optimization algorithms over graphs. *arXiv [Preprint]* arXiv:1704.01665. (2017).
29. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature*. (2015) 518:529–33. doi: 10.1038/nature14236
30. Wang B, Sun Y, Duong TQ, Nguyen LD, Hanzo L. Risk-aware identification of highly suspected covid-19 cases in social iot: a joint graph theory and reinforcement learning approach. *IEEE Access*. (2020) 8:115655–61. doi: 10.1109/ACCESS.2020.3003750
31. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
32. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-Sanitization: a semantic privacy-preserving framework for unstructured medical datasets. *Comput Commun*. (2020) 161:160–71. doi: 10.1016/j.comcom.2020.07.032
33. Iwendi C, Srivastava G, Khan S, Reddy Maddikunta PK. Cyberbullying detection solutions based on deep learning architectures. *Multimedia Syst*. (2020). doi: 10.1007/s00530-020-00701-5. [Epub ahead of print].
34. Manoj M, Srivastava G, Somayaji SRK, Gadekallu TR, Maddikunta PKR, Bhattacharya S. An incentive based approach for COVID-19 planning using

blockchain technology. In: *2020 IEEE Globecom Workshops*. GC Wkshps: IEEE (2020). p. 1–6. doi: 10.1109/GCWkshps50303.2020.9367469

35. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc*. (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
36. Schmidhuber J, Hochreiter S. Long short-term memory. *Neural Comput*. (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kumar, Khan, Din, Band, Mosavi and Ibeke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 11 October 2021
doi: 10.3389/fpubh.2021.751536

A Predictive and Preventive Model for Onset of Alzheimer's Disease

Udit Singhania[1], Balakrushna Tripathy[2], Mohammad Kamrul Hasan[3], Noble C. Anumbe[4], Dabiah Alboaneen[5], Fatima Rayan Awad Ahmed[6], Thowiba E. Ahmed[5] and Manasik M. Mohamed Nour[7]*

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, [2] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [3] Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia, [4] Department of Mechanical Engineering, University of South Carolina, Columbia, SC, United States, [5] Computer Science Department, College of Science and Humanities, Imam Abdulrahman Bin Faisal University, Jubail, Saudi Arabia, [6] Computer Science Department, College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, [7] Department of Mathematics, College of Science and Humanity, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Hikma Shabani,
Universiti Tun Hussein Onn Malaysia, Malaysia
Md. Sarwar Hosain,
Pabna University of Science & Technology, Bangladesh
Ebuka Ibeke,
Robert Gordon University, United Kingdom

***Correspondence:**
Mohammad Kamrul Hasan
mkhasan@ukm.edu.my;
hasankamrul@ieee.org

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *01 August 2021*
Accepted: *08 September 2021*
Published: *11 October 2021*

Citation:
Singhania U, Tripathy B, Hasan MK, Anumbe NC, Alboaneen D, Ahmed FRA, Ahmed TE and Nour MMM (2021) A Predictive and Preventive Model for Onset of Alzheimer's Disease. Front. Public Health 9:751536. doi: 10.3389/fpubh.2021.751536

Alzheimer's Disease (AD) is a neurodegenerative irreversible brain disorder that gradually wipes out the memory, thinking skills and eventually the ability to carry out day-to-day tasks. The amount of AD patients is rapidly increasing due to several lifestyle changes that affect biological functions. Detection of AD at its early stages helps in the treatment of patients. In this paper, a predictive and preventive model that uses biomarkers such as the amyloid-beta protein is proposed to detect, predict, and prevent AD onset. A Convolution Neural Network (CNN) based model is developed to predict AD at its early stages. The results obtained proved that the proposed model outperforms the traditional Machine Learning (ML) algorithms such as Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbor algorithms.

Keywords: machine learning, Alzheimer's disease, prediction, prevention, convolutional neural networks, support vector machine

INTRODUCTION

Alzheimer's disease (AD) is 60–70% the cause of dementia (1, 2). It is a slowly progressing brain disorder. The individuals that express symptoms of AD have abnormal deposition of a protein called amyloid-beta in their brain. This amyloid-beta protein forms plaques in the brain and strands of protein tau twists around, causing tangles that ultimately kills the brain cells. The degeneration of brain cells causes loss of memory, thinking and reasoning skills (3).

A few investigations have demonstrated that trained radiologists can be outflanked by computer-helped-diagnosis utilizing Support Vector Machines (SVMs) in distinguishing patients with AD and different ailments (4). Convolutional Neural Networks (CNNs) are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers (5). These are special type of Deep Neural Networks (DNNs) (6). Several applications of DNN can be found in (7). Deep learning techniques are used abundantly used in healthcare (8). For example various uses of CNNs for Magnetic Resonance Imaging (MRI) segmentation are presented in (9). CNNs are used for several applications Different groups have tried utilizing CNNs (10, 11) to analyze and separate AD from healthy or no condition (NC), patients using MRI scans as an input, while others have utilized a special type of MRI scan technique—the functional magnetic resonance imaging (fMRI) time-series

information (12, 13). An investigation was performed by Thompson et al. (14) in this regard in a paper titled, "Applying Convolutional Neural Networks for pre-detection of AD from structural MRI data." In this study, the authors utilized SVMs and CNNs on sectioned areas of interest after post-processing utilizing edge-detection algorithms on grouped MRI scan. This study achieved a sensitivity of 96% and a specificity of 98% on a dataset of 1,615 MRI images. Segmented images with edge-detection were used to classify them according to the required to the required quality standards.

The level of amyloid protein in the brain varies from individual to individual. Researchers have found (15, 16) that the rate of death from Alzheimer's has increased by 50% in the recent years, from 16 deaths in 1,00,000 in 1999 to 25 deaths in 1,00,000 in 2014. Also, the number of people who have died of AD has increased two-fold, from 44,536 in 1999 to 93,541 in 2014. In 2015, there were about 29.8 million people worldwide affected by AD. The cause of AD is still not properly understood. About 70% of the cause is inferred to be genetic in nature (17). A methodology to track, predict and cure AD can be analyzed with the history of the illness, cognitive testing with medical imaging and blood tests. People affected by AD chiefly rely on others for assistance which often turns out to be a burden on the caretakers. This results in social, psychological, physical, and economic pressure (18). Hence, we propose a model that would accept Positron Emission Tomography (PET) brain scans as inputs, process and analyze them and finally predict whether the individual will have Alzheimer's within a given time window of two years. It is said that prevention is better than cure. Hence, early detection of the onset of AD is highly beneficial. The techniques that are currently available to detect AD rely on cognitive impairment testing which is not efficient in yielding accurate diagnosis. The cognitive tests mostly rely on results of questionnaires directed toward the subjects and do not consider medical developments within the subject's brain explicitly. Such tests will prove to be an assuring method to predict the level of onset of AD in a subject only if there exists a one-on-one correspondence between the results of the tests and the physical state of the brain (the concentration of proteins that regulate the onset of AD). Thus, a new technique or algorithm was proposed in (19) with the aim of reducing the high dimensional MRI vector space to 150 dimensions using Principal Component Analysis (PCA). To categorize the reduced dimensions to PCA for progression of AD, multi-class neurons were employed. In comparative studies as in (20), Percent Whole Brain Volume Change (PBVC) was measured from serial MRI scans and dementia with Lewy bodies. The conclusion was that the atrophy of AD was significantly greater than that of Dementia with Lewy bodies (DLB) for one year in various regions of the brain including periventricular areas. PBVC was not significantly different from DLB and it was concluded that AD showed faster rate of global growth than DLB. Certain research has also been done to determine alterations occurring in Parahippocampal Cigumul bundle (PhC) and Posterior Cingulum bundle (PoC) in patients suffering from Mild Cognitive Impairment (MCI) (21, 22). This was done through diffusion tensor imaging (23–28). An atlas-based Region-Of-Interest (ROI) was used to calculate the fractional anisotropy, mean diffusivity, axial diffusivity, and radial diffusivity. For the primary health centre (PhC), a significant decrease was observed in the FA value, whereas an increase was observed in the MD and RD values. It is forecasted that by 2050, the prevalence of AD will quadruple to 26.6 million cases, where approximately 43% of them will need a high level of care. If the diagnosis of AD can be improved upon and treated at an earlier and more manageable stage, where therapeutics and preventive care are more useful, then the numbers of future AD patients will likely decrease.

The aim of this paper is to bring forward a predictive model for early prediction of AD and thus, develop a preventive model based on it. This model would be helpful in identifying patients with AD and those at risk of developing it. Furthermore, we suggest a preventive course of action to be followed to limit the growth in AD. In this paper, we propose an integrated algorithm that identifies a patient with AD and predicts the onset of AD with a data-oriented approach. In the predictive model, we consider amyloid protein concentration in MRI scans as one of the biomarkers to predict the onset of AD within a window of 2 years. Initially, based on MRI scans and cross sectional and longitudinal MRI scan datasets, we classify subjects as AD or No Condition (NC). Following this, we analyze the datasets to compute the accuracy of onset of AD for the subjects. This is without considering the subjects classified as AD or NC by the first algorithm. We further use this to outline the preventive measures that should be taken to prevent (or delay) the onset or worsening of AD. The preventive measures are aimed at reducing the Critical Design Reviews (CDR) score and hence, reduce the severity of the onset of AD. The rest of the paper is organized as follows: In materials and methods, we broadly discuss the algorithm that we suggest with the data used, the model description and the performance of the algorithm. In results and analysis, we analyze the results of our algorithm and discuss the same. In result analysis, we provided a conclusion that briefly describes results and analysis and provides further improvement and upgrades that can be made in the proposed algorithm.

MATERIALS AND METHODS

In this section, we broadly discuss the architectural design of the algorithm, the types of data used and the methods of data acquisition. We also discuss the performance of the algorithm and the classification algorithms used.

Subjects and Data Acquisition

We have taken a data-oriented approach while developing the algorithm. This algorithm is developed consulting the existing algorithms. The data used for the algorithm was obtained from an Operational Applications of Special Intelligence Systems (29–32). It consists of cross-sectional MRI scan collection of 416 individuals within the age group of 18 to 96. About 100 of the individuals mentioned in the dataset were clinically diagnosed with mild or moderate AD. The data used here also comprises of longitudinal MRI scan collection of 373 individuals within the age group of 18 to 96. The dataset consisted of biomarkers

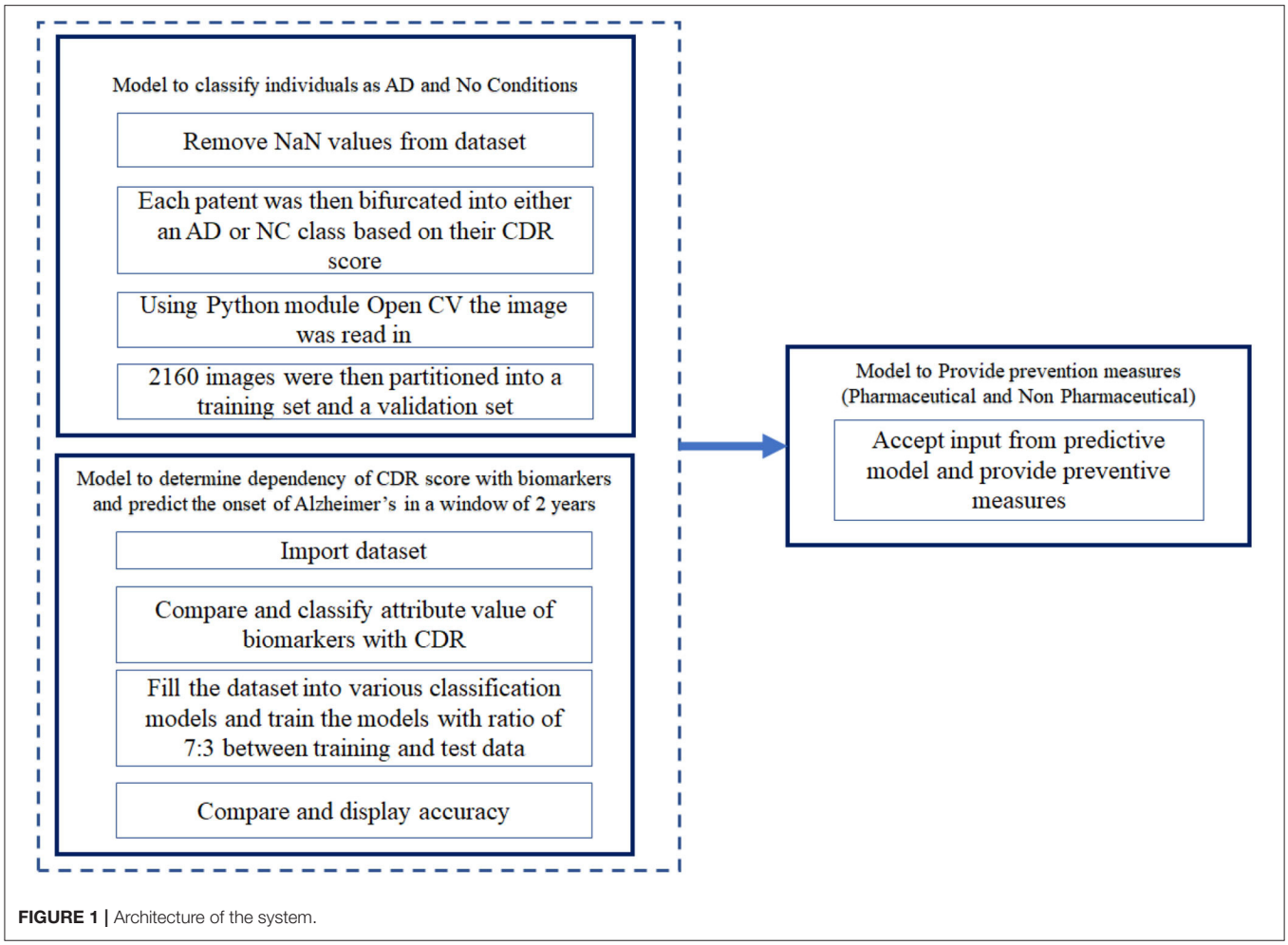


FIGURE 1 | Architecture of the system.

that have been scientifically proven to predict onset of AD. The primary attributes of the dataset were the Minimal Mental State Examination (MMSE) scores and the Clinical Dementia Rating (CDR) scores.

Model Description

The algorithm developed can be visualized as two modules, namely, the predictive model and the preventive model. The predictive model takes into consideration the dataset mentioned above and uses it to predict the possibility of onset of AD (**Figure 1**). The predictive model is divided into two sub models. The first model of the predictive algorithm takes into consideration the cross sectional dataset and the MRI scan images of the subjects mentioned in the cross sectional dataset and confirms whether a given subject is affected by AD currently. Initially, it removes null values from the dataset by dropping rows that have one such value. Due to unavailability of diverse range of Critical Design Reviews (CDR) scores, any individual having a CDR score equal to 0 was considered to be Non-Cognizable (NC) and the remaining CDR scores implied AD. Now, the removal of rows consisting of "NaN" values reduced the datasets to 216 data points and each of the individuals had 10 coronal slices of MRI scans.

Using the OpenCV algorithm available, the images were read as a numpy array. This was then normalized by dividing each intensity by 255, which is the maximum intensity. The k-means algorithm was run on the pixel space of each image. In the analysis phase, only two clusters were chosen for simplicity and maximum noise reduction. Each pixel was then assigned to the cluster it was in. The 2,160 image set obtained was divided into training set and validation set, such that a ratio of 7:3 was maintained between the training and validation set. The training set was then fed into the CNN, VGG16 in batches of 32. The model flattens the output and feeds it into two fully-connected or dense layers each containing 1,028 connecting units. Using a softmax activation function in the final layer, stochastic gradient descent as an optimizer, categorical cross entropy as a loss, the model was trained for 150 epochs. This algorithm classifies individuals with present traits of AD. In the second model of the predictive algorithm, the cross sectional and longitudinal datasets were fit into various classification models such as Decision Tree Classifier, Logistic Regression, TensorFlow, K nearest neighbour (KNN) (33–35), SVM etc. and compare the accuracy with which these classification algorithms predict the CDR scores when

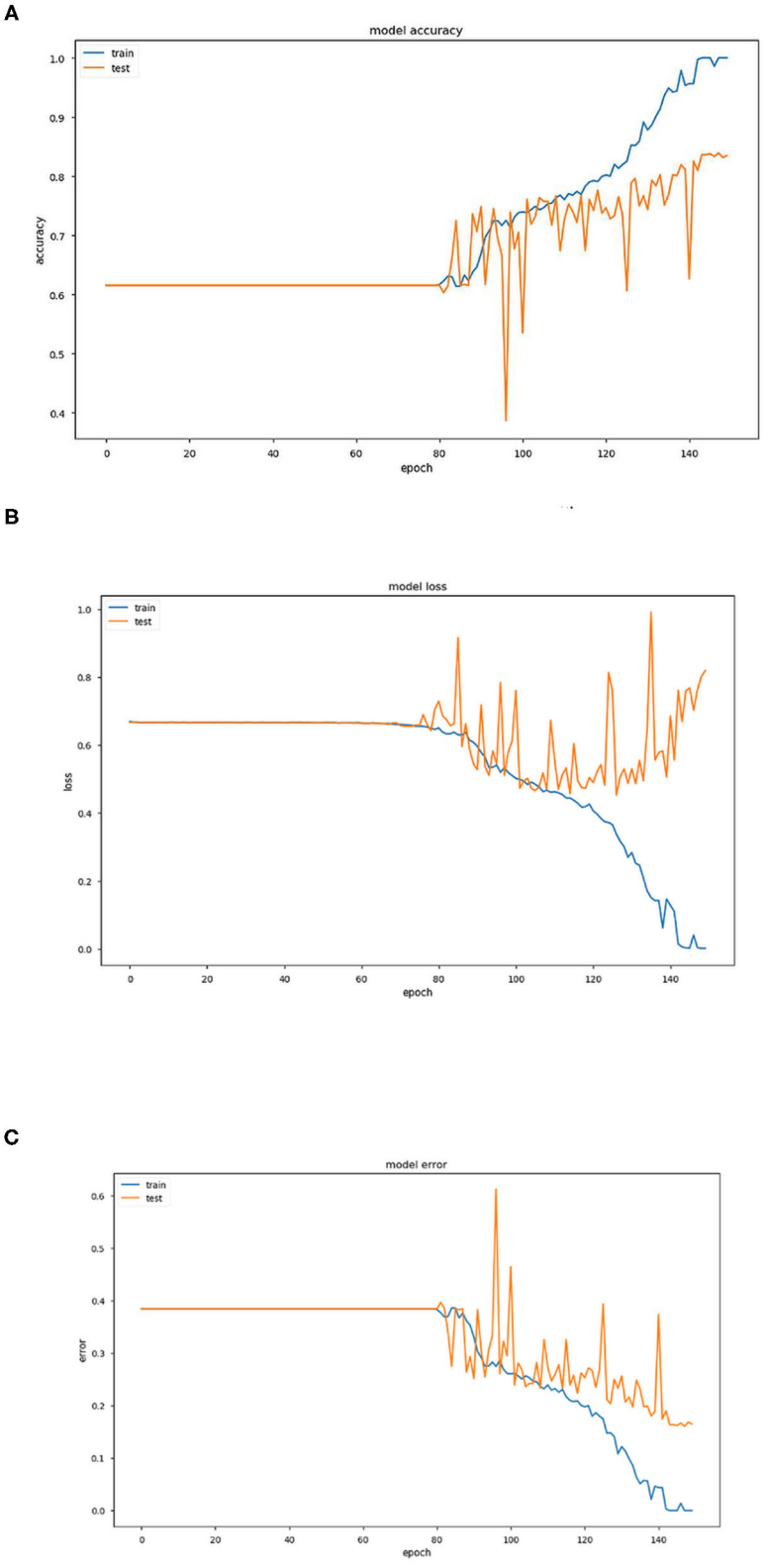


FIGURE 2 | (A) Graph representing variation of model accuracy with respect to epoch. **(B)** Graph representing variation of model loss with respect to epoch. **(C)** Graph representing variation of model error with respect to epoch.

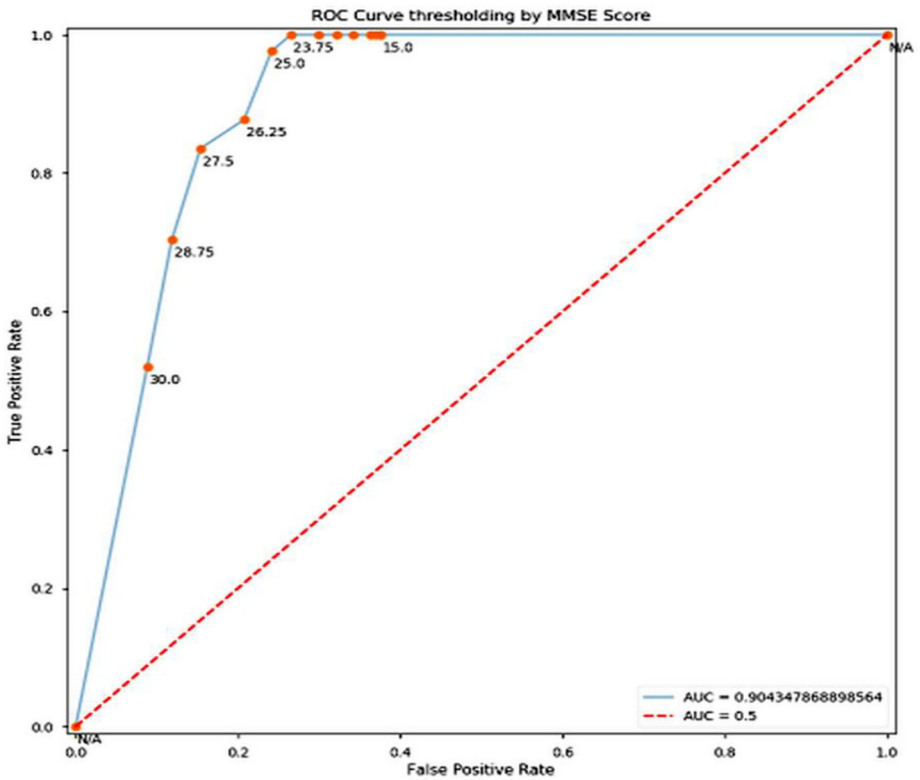


FIGURE 3 | ROC Curve.

provided the set of biomarkers and parameters available in the dataset. In the preventive model, we consider the most recent CDR score of the subject and accordingly provide the clinically approved preventive measure.

Performance Measurement and Classification

For a fixed set of Access Points (Aps), the probability of a user to be closer to any three out of the four APs is greater than that of being at equal distances from each of the APs. At every user location, the updated list of three APs based on maximum RSSI and minimum Program Visualization (PV) distances are utilized in the *k*-NN search to get the least localization error. Maximization of the RSSI objective function is computed as;

$$f(x) = B(0) + \sum_{i=1}^{n} \left(a_i * (x, x_i)\right) \quad (1)$$

Here, x represents the new input vector and a_i, x_i represents all the weights of the neural and the support vectors obtained from the training data respectively. *B (0)* is the bias input chosen. On the other hand, logistic regression is used for the binary classification and prediction using the parameters and biomarkers mentioned above to classify the subjects as AD or NC based on the MMSE and CDR scores available in the training data. The logistic function used to obtain the predicted value is given by:

$$f(x) = B(0) + \sum_{i=1}^{n} e^{(b_0 + b_i * x)} / (1 + e^{(b_0 + b_i * x)}) \quad (2)$$

Where b_0 represents the bias used in the network and b_i represents the vector values obtained from the training data. There were other classification algorithms used to predict the CDR scores. The architecture of the proposed system is depicted in **Figure 1**.

RESULTS AND ANALYSIS

In this section, we mention the results of the algorithm and discuss the same. The experimental evaluation approach has considered for the assessment of the proposed method. The algorithms were run on Intel(R) Core (TM) i5-7200U CPU with 2.50GHz. The experimental result is presented in **Figures 2**, **3**.

The above three **Figures 2A–C** provide estimations of the model accuracy, model loss and model error with respect to the increase in epoch during the training phase and testing phase. Before epoch reaches 80, all the three parameters of accuracy, loss and error remain constant and also coincide. The variations occur after this in the training phase and the accuracy increases continuously, the loss and error decrease as expected. In the

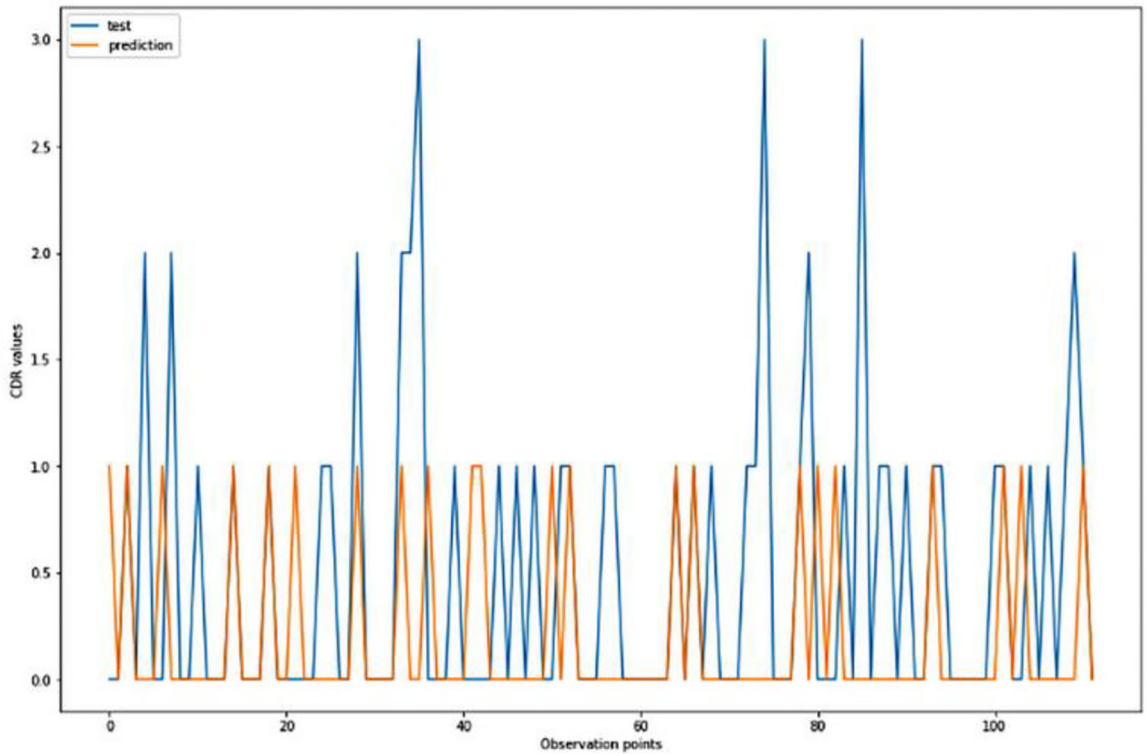


FIGURE 4 | Graphical representation of the comparison of test and predicted CDR values obtained using the logistic regression algorithm. (six parameters).

testing phase also, the same trend is observed but the extent becomes much restricted with large variations.

Training and test accuracy were measured at each epoch. The model stagnated and predicted that all brain scans were NC for the first 81 epochs. Then it moved out of its local minima and began increasing its accuracy. The training error ultimately became 0% at epoch 144, thus showing it had predicted all the brains correctly. This can be taken as an indication of overfitting on the training set. However, the lowest prediction error reached is 16.05%, by epoch 148. It can be argued that there could be a further decrease in the prediction error. But due to the model reaching 0% training error, it was stopped prematurely. As a consequence a part of the training dataset is typically set aside as the "test set" to check for overfitting.

Predictive Model

For the algorithm that predicts whether a person has AD or not, each epoch took around 1,890 s to train. The total number of epochs run was 150, which took about 4,725 min or around 79 h.

Training and test accuracy were measured at each epoch. The model stagnated and predicted that all brain scans were NC for the first 81 epochs. Then it moved out of its local minima and began increasing its accuracy. The training error ultimately became 0% at epoch 144, thus showing it had predicted all the brains correctly. This can be taken as an indication of overfitting on the training set. However, the lowest prediction error reached is 16.05%, by epoch 148. It can be argued that there could be a further decrease in the prediction error. But due to the model reaching 0% training error, it was stopped prematurely An ROC (Receiver Operating Characteristic) Curve was plotted by thresholding against a patient's Minimum Mean-Square Error (MMSE) scores, to evaluate the model (**Figure 3**). The ROC is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

In conclusion, it can be stated that this model of pre-processing with k-means and training with VGG16 performed well (training accuracy = 100%, prediction accuracy = 83.95%, AUC = 0.904). It can perform better by increasing the number of clusters in the k-means step, combining edge-detection and segmentation steps, using better hardware and larger networks in terms of the number of layers and epochs run, and in general obtaining of more data. In terms of engineering the biological system of the brain in the context of AD, this model does well in interpreting the brain as a set of pixels and their intensities but easily disregards the numerous other variables that assemble a diagnosis, especially considering that MRI scans are already a large abstraction from the vastly complex human brain.

For the algorithm that predicts the onset (or worsening) of AD, the biomarkers and parameters were considered against the CDR score presented in **Figure 4**. The data was divided with a ratio of 7:3 and then fit into various classification models to predict the CDR scores. It is widely known that The CDR scores of 0 means "Normal," 0.5 means "Very Mild Dementia," 1 means "Mild Dementia," 2 means "Moderate Dementia" and 3 means

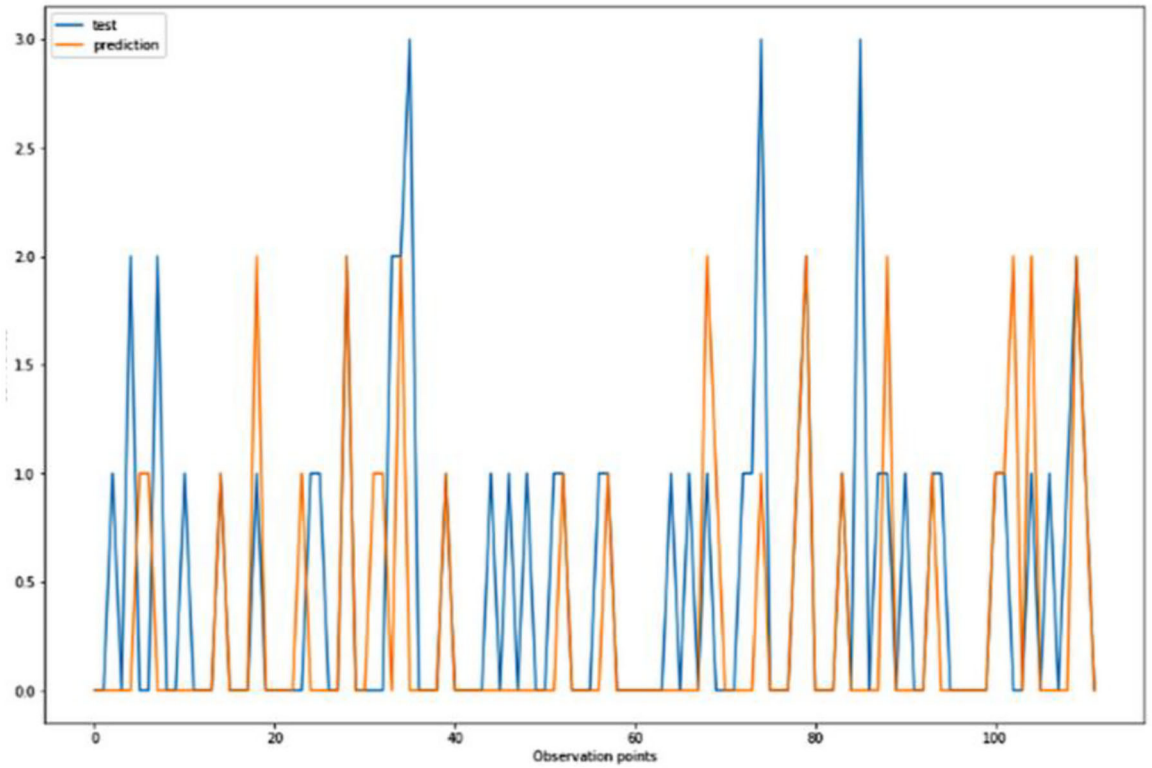


FIGURE 5 | Graphical representation of the comparison of test and predicted CDR values obtained using the k-neighbors classification algorithm. (six parameters).

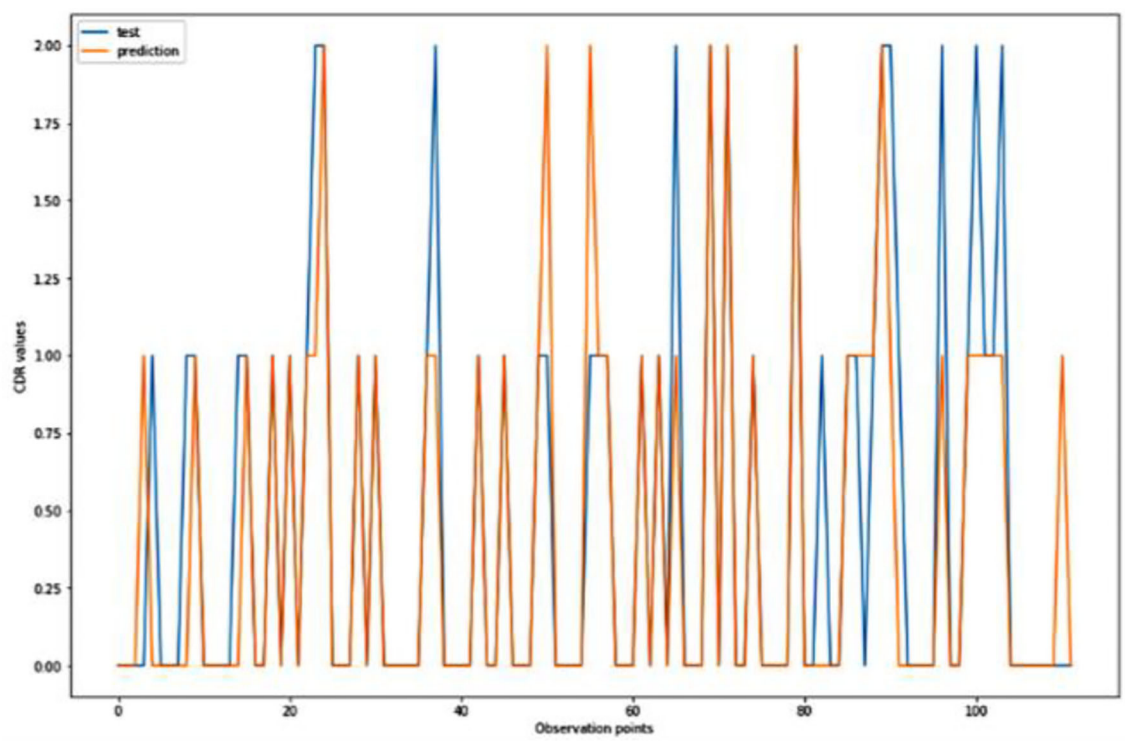


FIGURE 6 | Graphical representation of the comparison of test and predicted CDR values obtained using the logistic algorithm. (11 parameters).

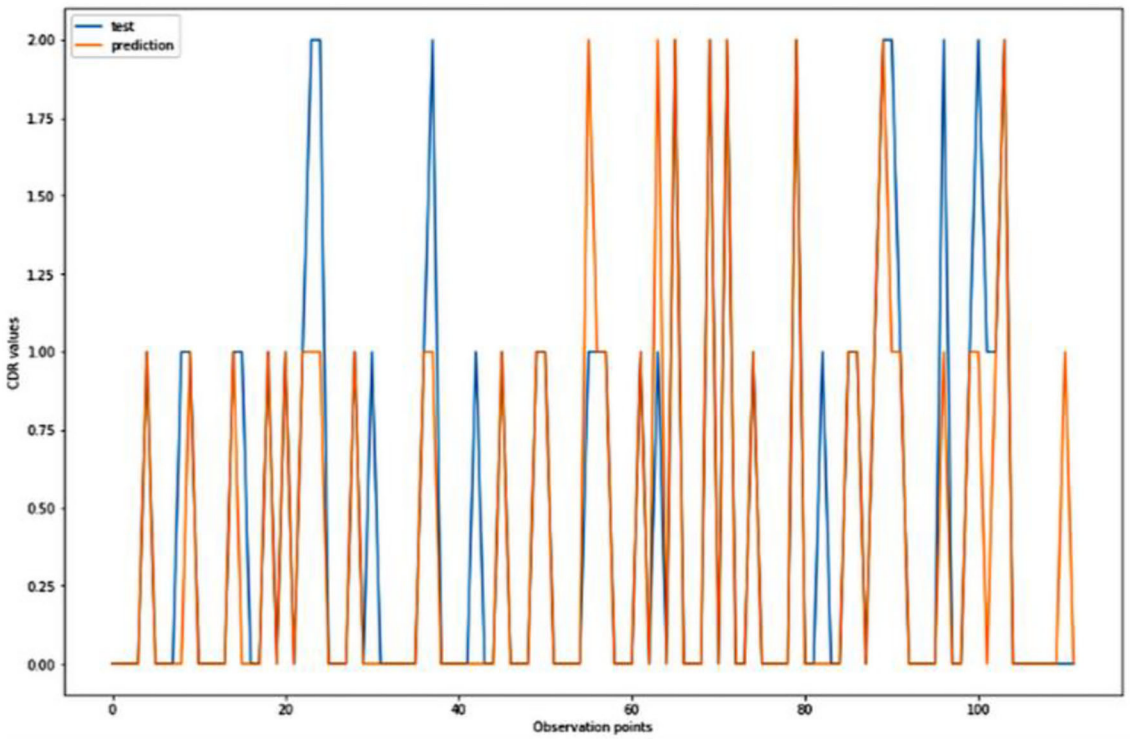


FIGURE 7 | Graphical representation of the comparison of test and predicted CDR values obtained using the k-neighbors classification algorithm. (11 parameters).

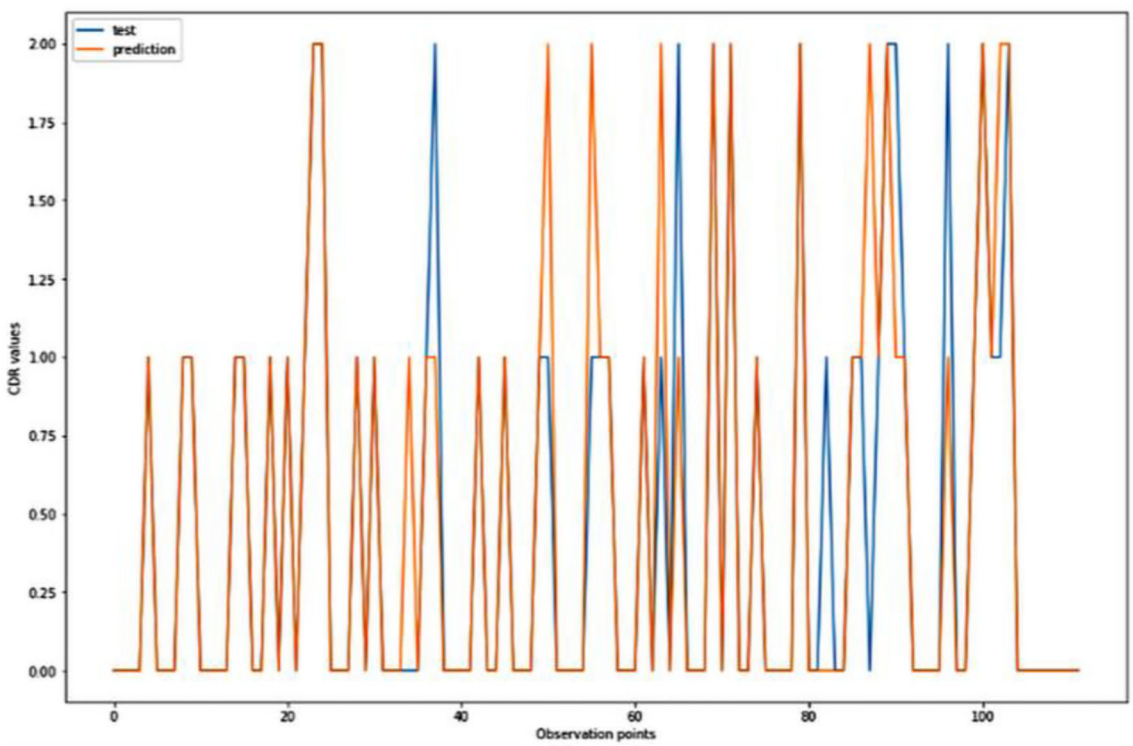


FIGURE 8 | Graphical representation of the comparison of test and predicted CDR values obtained using the TensorFlow model. (11 parameters).

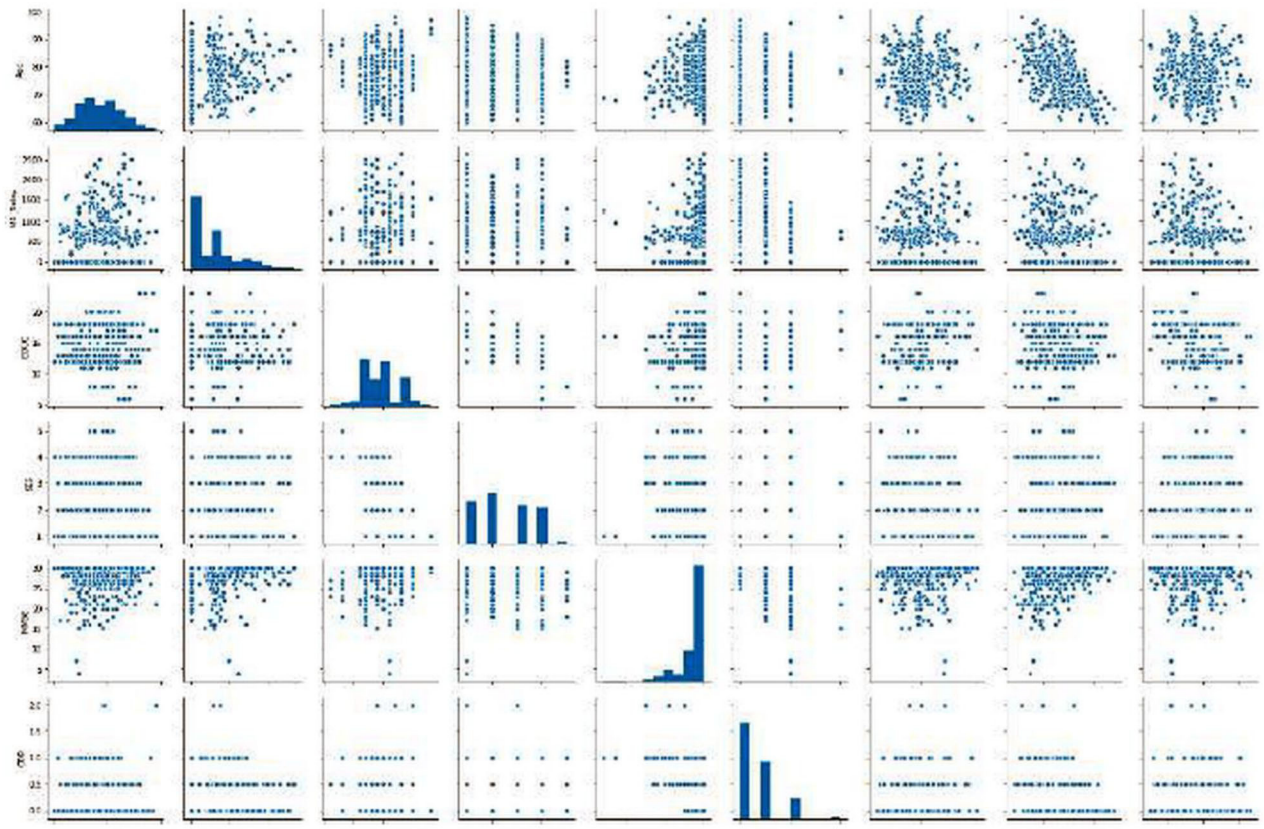


FIGURE 9 | Graphics of scatter plots representing the correlation matrix of individual parameters against the rest of the parameters in consideration.

"Severe Dementia". Accordingly, the results in the **Figures 4–8** are interpreted for test and prediction. Further interpretations can be obtained from the table (36).

Figures 4–8 presented the CDR using the different parameters. Here, "sklearn" was used as a metric, whereas "TensorFlow" was used as a deep neural network classifier. There are two sets of results obtained—one by training the algorithm with the longitudinal and cross-sectional datasets, the other because of concatenating the datasets. The results obtained on training the neural network with the cross sectional and longitudinal datasets varied extensively for different models of classification. Here, results refer to the percentage of accuracy with which the model predicts onset of dementia or AD. The datasets were split into training data and test data using the CDR model with the training data: test data ration equal to 7:3. For the longitudinal dataset, the training data successfully predicted the onset of Alzheimer's with an accuracy of 83–88% for different classification models such as K Nearest Neighbors, SVM, TensorFlow, Standard Scalar and Logistic Regression, whereas the test data predicted the same result with an accuracy of 60–70%. The graphical representation of the correlation matrix proved that the dataset is random in nature and does not have any hidden pattern in it. On the other hand, when the datasets were classified using the CDR and MMSE scores and comparing them with the MRI scans, such that a ratio of 7:3 is maintained between the training data and test data set populations, the training data, and the test data both predicted the onset of AD with an accuracy of ∼85%.

Different scatter plots (**Figure 9**) are also plotted which represent the correlation matrix of individual parameters against the rest of the parameters in consideration. Different scatter plots (**Figure 9**) are also plotted which represent the correlation matrix of individual parameters against the rest of the parameters in consideration. Summarized results are presented in **Tables 1**, **2**.

Preventive Model

We further implemented a system which would display the preventive measures that should be taken by a patient (patient ID is provided as input) to prevent onset or control the severity of AD. The measures given for a particular patient were according to the present condition of the patient and keeping in mind that the condition will worsen (since AD is degenerative) over time if preventive measures are not taken. Any medication, if and wherever suggested were in accordance with the U.S. Food and Drug Administration (FDA). Also, medications were only suggested for the patients who are currently suffering from AD.

TABLE 1 | Accuracy obtained to predict AD in 2 years on the training and test sets.

Models	Accuracy			
	six Parameters (M/F, Age, EDUC, SES, eTIV, ASF)		11 Parameters (M/F, Age, EDUC, SES, eTIV, ASF, Group, Hand, MMSE, MR Delay, nWBV)	
	Training Set	Test Set	Training Set	Test Set
Logistic Regression	59.00%	67.86%	83.14%	78.57%
Decision Tree Classifier	98.47%	72.32%	95.40%	86.61%
K Nearest Neighbor	86.21%	68.75%	90.42%	79.46%
Support Vector Classifier	52.87%	60.71%	–	–

TABLE 2 | Prediction accuracy for AD.

Models	Prediction Accuracy	
	six Parameters (M/F, Age, EDUC, SES, eTIV, ASF)	11 Parameters (M/F, Age, EDUC, SES, eTIV, ASF, Group, Hand, MMSE, MR Delay, nWBV)
Sklearn	70.54%	84.82%
TensorFlow	70.54%	84.82%

RESULT ANALYSIS

The predictive and preventive models were implemented successfully. The predictive model consists of two parts: to predict This model can further be improved upon by predicting the individuals who could develop AD or whose condition can worsen, instead of predicting for the whole dataset. Moreover, including genetic mappings, psychometric tests, mini mental state examination tests along with hand drawn images and shapes, etc. can make this whole system more comprehensive for the prediction of Alzheimer's disease. Systems with higher specifications can reduce the execution time. The algorithm can be further optimized and generalized using fuzzy logic concepts, such as fuzzy c-means instead of k-means algorithm which is essentially binary logic concept as implemented in this algorithm. Furthermore, hybrid models for clustering algorithms can also be used. One of the important prospects of this algorithm is to identify whether the preventive measures influence the subjects, either with mild, early stage or severe AD and to analyze the same based on MMSE and CDR scores over a period. Finally, the model can also be tested for scalability on big data (25, 30, 36–38). Whether a person has AD or not, using the T1 weighted coronal brain scan MRI, and to predict the severity of AD in the next 2 years, using the data from processed MRI images and through amyloid protein concentration. This achieved an accuracy of almost 85% and can be further improved as the number of executions is increased and data is added.

CONCLUSIONS

The preventive model was based on the present condition of the patient and the predictive model. The preventive measures for each patient can be obtained by giving the ID of the patient as an input. With the current treatments of AD only tackling the symptoms, research is needed to check the onset of AD along with a way to deal with the biological changes that are responsible for it. With more data on the patients of AD (or suspected) the algorithms can be improved upon. As is evident, more parameters improve the accuracy of the algorithm hence; research on more biomarkers that play a role in the progression of AD is required.

DATA AVAILABILITY STATEMENT

Data were provided by OASIS Cross-Sectional and Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J, Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, and U24 RR021382.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

This research has been supported by the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM) under the grant GGPM 2020-028, and Prince Sattam Bin Abdulaziz University, Saudi Arabia. Data were provided by OASIS Cross-Sectional and Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J, Csernansky J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, and U24 RR021382.

REFERENCES

1. Burns A, Iliffe S. Alzheimer's disease. *BMJ.* (2009) 338:b158. doi: 10.1136/bmj.b158
2. World Health Organization. Fact sheet Number 362 Dementia. Available www.who.int/medicentre/factsheets/fs363/en/
3. O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci.* (2011) 34:185–204. doi: 10.1146/annurev-neuro-061010-113613
4. López MM, Ramírez J, Górriz JM, Álvarez I, Salas-Gonzalez D. Segovia, F, et al., SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. *Neurosci Lett.* (2019) 464:233–8. doi: 10.1016/j.neulet.2009.08.061
5. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, et al. Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front Neurosci.* (2018) 12:777. doi: 10.3389/fnins.2018.00777
6. Zeng N, Qiu H, Wang Z, Liu W, Zhang H, Li Y, et al. new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing.* (2018) 320:195–202. doi: 10.1016/j.neucom.2018.09.001
7. Klöppel S, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, et al. Accuracy of dementia diagnosis—a direct comparison between radiologists and a computerized method. *Brain.* (2008) 131:2969–74. doi: 10.1093/brain/awn239
8. Payan A, Montana G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arxiv[preprint].arxiv:1502.02506, 2015.
9. Mathotaarachchi S, Pascoal TA, Shin M, Benedet L, Kang MS, Beaudry T, et al. Alzheimer's disease neuroimaging initiative-identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiol Aging.* (2017) 59:80–90. doi: 10.1016/j.neurobiolaging.2017.06.027
10. Sarraf S and Tofighi G. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. arxiv[preprint].arxiv:1603.08631, 2016.
11. Gunawardena, KANNP, Rajapakse RN and Kodikara ND. "Applying convolutional neural networks for pre-detection of alzheimer's disease from structural MRI data," In: *24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP).* Auckland (2017). p. 1-7. doi: 10.1109/M2VIP.2017.8211486
12. Taylor CA, Greenlund SF, McGuire LC, Lu H. and Croft JB. Deaths from Alzheimer's disease—United States. *Morb Mortal Wkly Rep.* (2017) 66:1999–2014. doi: 10.15585/mmwr.mm6620a1
13. Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E. Alzheimer's disease. *Lancet.* (2011) 377:1019–31. doi: 10.1016/S0140-6736(10)61349-9
14. Thompson CA, Spilsbury K, Hall J, Birks Y, Barnes Y, Adamson J. Systematic review of information and support interventions for caregivers of people with dementia. *BMC Geriatr.* (2007) 7:1–18. doi: 10.1186/1471-2318-7-18
15. Mahmood R, and Ghimire B. Automatic detection and classification of Alzheimer's Disease from MRI scans using principal component analysis and artificial neural networks," In: *20th International Conference on Systems, Signals and Image Processing (IWSSIP).* Bucharest (2013), p. 133-137. doi: 10.1109/IWSSIP.2013.6623471
16. Mak E, Su L, Williams GB, Watson R, Firbank M, Blamire AM, et al. Longitudinal assessment of global and regional atrophy rates in Alzheimer's disease and dementia with Lewy bodies. *Neuroimage Clin.* (2015) 7:456–62. doi: 10.1016/j.nicl.2015.01.017
17. Brookmeyer R, Johnson E, Ziegler-Graham K. and Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimer's Dement.* (2007) 3:186–91. doi: 10.1016/j.jalz.2007.04.381
18. Bubb EJ, Metzler-Baddeley C. and Aggleton JP. The cingulum bundle: anatomy, function, and dysfunction. *Neurosci Biobehav Rev.* (2018) 92:104–27. doi: 10.1016/j.neubiorev.2018.05.008
19. Urger E, DeBellis MD, Hooper SR, Woolley DP, Chen S, Provenzale JM. Influence of analysis technique on measurement of diffusion tensor imaging parameters. *Am J Roentgenol.* (2013) 200:W510–7. doi: 10.2214/AJR.12.9650
20. Marcus, DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J Cogn Neurosci.* (2007) 19:1498-507. doi: 10.1162/jocn.2007.19.9.1498
21. Marcus, DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J Cogn Neurosci.* (2010) 22:2677–84. doi: 10.1162/jocn.2009.21407
22. Patel H, Singh Rajput D, Thippa Reddy G, Iwendi C, Kashif Bashir A, Jo O, et al. review on classification of imbalanced data for wireless sensor networks. *Int J Distrib Sens Netw.* (2020) 16:1–15. doi: 10.1177/1550147720916404
23. Tripathy BK, Parimala M. and Reddy GT. Innovative classification, regression model for predicting various diseases. In: Lee KC, Roy SS, Samui P, Kumar V, editors. *Data Analytics in Biomedical Engineering and Healthcare.* Academic Press (2020). p. 179–203. doi: 10.1016/B978-0-12-819314-3.00012-4
24. Tripathy BK, Mohanty RK and Sooraj TR. "On intuitionistic fuzzy soft set and its application in group decision making," In: *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETEk.TS).* Pudukkottai (2016). doi: 10.1109/ICETETS.2016.7603002
25. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HM, et al. Hep-Pred: Hepatitis C staging prediction using fine gaussian SVM. *Comput Mater Continua.* 69:191-203. doi: 10.32604/cmc.2021.015436
26. Hasan MK, Shafiq M, Islam S, Pandey B, Baker El-Ebiary YA, Nafi NS, et al. Lightweight cryptographic algorithms for guessing attack protection in complex internet of things applications. *Complexity.* (2021) 2021:5540296. doi: 10.1155/2021/5540296
27. Hasan MK, Islam S, Sulaiman R, Khan S, Hashim AH, Habib S, et al. Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access.* (2021) 9:47731-42. doi: 10.1109/ACCESS.2021.3061710
28. Hasan MK, Ahmed MM, Musa SS, Islam S, Abdullah SNHS, Hossain E, et al. An improved dynamic thermal current rating model for PMU-based wide area measurement framework for reliability analysis utilizing sensor cloud system. *IEEE Access.* (2021) 9:14446-58. doi: 10.1109/ACCESS.2021.3052368
29. Tripathy BK. and Nanda S. Absolute value of fuzzy real numbers and fuzzy sequence spaces. *J Fuzzy Math.* (2000) 8:883–92.
30. Iwendi C, Khan S, Anajemba JH, Bashir AK. and Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access.* (2020) 8:28462–74. doi: 10.1109/ACCESS.2020.2968537
31. Nahla N, Kamrul Hasan M, Memon I, Saeed RA, Ariffin KAZ, Ali ES, et al. A systematic review on cognitive radio in low power wide area network for industrial IoT applications. *Sustainability.* (2021) 13:338. doi: 10.3390/su13010338
32. Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, et al. IoT for Smart Cities: Machine learning approaches in smart healthcare—a review. *Future Internet.* (2021) 13:218. doi: 10.3390/fi13080218
33. Memon I, Shaikh RA, Hasan MK, Hassan R, Haq AU, Zainol KA. Protect mobile travelers information in sensitive region based on fuzzy logic in IoT technology. *Secur Commun Netw.* (2020) 2020:8897098. doi: 10.1155/2020/8897098
34. Meri A, Hasa MK. and Safie N. Success factors affecting the healthcare professionals to utilize cloud computing services. *Asia Pac J Inf Technol Multimedia.* (2017) 6:31–42. doi: 10.17576/apjitm-2017-0602-04
35. Ghazal TM, Hussain MZ, Said RA, Nadeem A, Hasan MK, Ahmad M, et al. Performances of K-means clustering algorithm with different distance metrics. *Intell Auto Soft Comput.* (2021) 30:735–742. doi: 10.32604/iasc.2021.019067
36. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Reddy Maddikunta PK, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics.* (2020) 9:1–16. doi: 10.3390/electronics9020274
37. Reddy GT, Reddy MPK, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G, et al. Analysis of dimensionality reduction techniques on big data. *IEEE Access.* (2020) 8:54776–88. doi: 10.1109/ACCESS.2020.2980942
38. Deepa N, Pham QV, Nguyen DC, Bhattacharya S, Gadekallu TR, Reddy Maddikunta PK et al. A Survey on blockchain for big data:

approaches, opportunities, and future directions. arxiv[preprint].arxiv:2009.00858, 2020.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Singhania, Tripathy, Hasan, Anumbe, Alboaneen, Ahmed, Ahmed and Nour. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 12 October 2021
doi: 10.3389/fpubh.2021.737149

Fischer Linear Discrimination and Quadratic Discrimination Analysis–Based Data Mining Technique for Internet of Things Framework for Healthcare

Mohammad Kamrul Hasan[1], Taher M. Ghazal[1,2], Ali Alkhalifah[3], Khairul Azmi Abu Bakar[1], Alireza Omidvar[4], Nazmus S. Nafi[5] and Johnson I. Agbinya[5]*

[1] Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia, [2] Skyline University College, University City of Sharjah, Sharjah, United Arab Emirates, [3] Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia, [4] Engineering Department, Tehran Urban and Suburban Railway Co., Tehran, Iran, [5] School of IT and Engineering (SITE), Melbourne Institute of Technology, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Aruna M. G., Visvesvaraya Technological University, India
Areej Fatima, Lahore Garrison University, Pakistan
Bishwajeet Pandey, Gran Sasso Science Institute, Italy

***Correspondence:**
Mohammad Kamrul Hasan
mkhasan@ukm.edu.my

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *06 July 2021*
Accepted: *20 August 2021*
Published: *12 October 2021*

Citation:
Hasan MK, Ghazal TM, Alkhalifah A, Abu Bakar KA, Omidvar A, Nafi NS and Agbinya JI (2021) Fischer Linear Discrimination and Quadratic Discrimination Analysis–Based Data Mining Technique for Internet of Things Framework for Healthcare. Front. Public Health 9:737149. doi: 10.3389/fpubh.2021.737149

The internet of reality or augmented reality has been considered a breakthrough and an outstanding critical mutation with an emphasis on data mining leading to dismantling of some of its assumptions among several of its stakeholders. In this work, we study the pillars of these technologies connected to web usage as the Internet of things (IoT) system's healthcare infrastructure. We used several data mining techniques to evaluate the online advertisement data set, which can be categorized as high dimensional with 1,553 attributes, and the imbalanced data set, which automatically simulates an IoT discrimination problem. The proposed methodology applies Fischer linear discrimination analysis (FLDA) and quadratic discrimination analysis (QDA) within random projection (RP) filters to compare our runtime and accuracy with support vector machine (SVM), K-nearest neighbor (KNN), and Multilayer perceptron (MLP) in IoT-based systems. Finally, the impact on number of projections was practically experimented, and the sensitivity of both FLDA and QDA with regard to precision and runtime was found to be challenging. The modeling results show not only improved accuracy, but also runtime improvements. When compared with SVM, KNN, and MLP in QDA and FLDA, runtime shortens by 20 times in our chosen data set simulated for a healthcare framework. The RP filtering in the preprocessing stage of the attribute selection, fulfilling the model's runtime, is a standpoint in the IoT industry.

Index Terms: Data Mining, Random Projection, Fischer Linear Discriminant Analysis, Online Advertisement Dataset, Quadratic Discriminant Analysis, Feature Selection, Internet of Things.

Keywords: data mining, Fischer linear discriminant analysis, quadratic discriminant analysis (QDA), feature exaction and selection, Internet of Things—IoT, healthcare applications

INTRODUCTION

The importance of information technology in the age of communication is not hidden from anybody. In 2021, more than 35 billion things are connected, and lots of potentials can be utilized in Internet of things (IoT) healthcare applications. It is not unrealistic to conceive that in the near future, artificial intelligence can substitute the service-based evaluators and manipulate IoTs that are much more accurate, professional, and appealing instead of ones with various privacy gaps. They can be accessed through a form of robots that hijack their information, resulting in incorrect information circulation and also increasing the riskiness of information of healthcare services. The third wave Web x.0 is now evolving by enhancing the Internet capacity in which companies today use the fourth generation of the IoT for their enterprise resource planning (ERP) applications. This generation has reached cloud computing, multiplying its capabilities in the virtual world in terms of velocity, veracity, and variability in crowdsourcing projects, such as grid.to, or other crowdsourcing events in cyberspace, which has achieved a high synergy in terms of availability and responsiveness. The best learned lessons on services such as Uber—which already managed to attract millions of visitors—was the reason that motivated us to perform a classification algorithm on this data set of web usage in which many points of view, such as the ethical, professional, civilian's rights, and media laws, are likewise healthcare data sets and should be taken into account.

Online advertisements in widgets, banners, and plugins are always being updated by the service providers. Their components are now coded by XML[1] or AJAX[2] compilers acting on the service side, leading to optimization of the server idle time. These web services are often programmed in PHP or other frameworks, such as Perl or CSS. But the design, maintenance, and updating of such services are necessarily done as separately managed projects, in which the result is aggregated in the service according to the perception of end users and other stakeholders. For instance, the Bitcoin exchange intermediaries generally compete in a win–win or offensive strategy for advertising luxury brands, soccer competitions through reputable sponsoring channels, and specialty stores and branches on free trade zones to enter the market decentral delivering the service at any location. Here, Bitcoin miners and exchangers are thought to be among the most critical stakeholders in the affluent class. Their emotional behavior is accompanied by their high expected reliability and tangibility of the service at very high prices. During the operation of Digital Wallet, many stakeholders decide to research on the queuing systems (1), maintain their privacy on attacker behavior, or change their service's ERP system by taking advantage of a new big data system on security concerns (2, 3).

We have selected this standpoint for the selection of our data set. Therefore, in this research, we presume that the web usage features are among the most important topics in the scope of the IoT and would be studied especially for healthcare frameworks with machine learning.

In this research, the ultimate goal is to study the dynamics of the IoT-intensive systems for data mining resources, such as runtime and responsiveness, and also in the essential part of analysis regarding the transparency of structure and reliability measured by accuracy. Therefore, analysis of modeling focuses on the accuracy of Internet advertisement recognition and relies on the runtime on the unstructured high-volume data as an example of IoT systems. In this paper, we first review the literature in the field of web usage and data mining. Then, the ethical and professional aspects are discussed to link the framework to practice. Finally, the results of the modeling are represented and compared with recent works in this area. The work outlook is discussed in the conclusion, in which the proposed methodology challenges can be aggregated to IoT evaluation frameworks such as service empathy, user-friendliness, responsibility, trust, and tangibility. This leads especially to achieve the solution for startups in the IoT industry in the legal, social, and even more professional sectors, such as health.

ETHICAL AND PROFESSIONAL ISSUES IN IoT

Nowadays, online user experience has emerged among healthcare and IoT systems, demonstrating a set of secure feature locations to other applications (4). These links could be another application or a social network that is a place to discuss and critique the product or a link to a vendor partner or a public health information panel.

How do different brands in the advertising industry measure their effectiveness in online advertising and their limitations in this way, and how do they overcome it? Principally, the answer to this question depends on customer feedback. This is because customer feedback results from the customer interaction with the service through an omnichannel. Google is using web analytics widgets to display complete information about web historiography. The same widgets for details about corporate websites include a list of links to a specific brand's website announcing some related keywords (5).

In an online advertisement, another thing that brand owners consider is the emotional aspects of the customer. The customer's memories, feelings, and mood when he or she received the product or service make him or her buy again. A dynamic co-creation framework is recently developed to measure trust as a good advertising motive for adding value (6). The trust can be viewed from two dimensions, namely, the dimensions of affectional and behavioral trust. Affections are related to emotions and the atmosphere in which the service is experienced, and consumer behavior is subjective and is specific to the personality. Affectional trust has opened a new way to effectively connect customer feeling to the space in which advertisement is taking place. This can effectively be implemented in the IoT systems when the users trust the quality of service in that framework. Establishing a relationship with the customer

[1]Extended Mark-up Language.
[2]Asynchronous Java-scripts.

can be inspired by new ideas and can be considered an advantage. This connection plays a significant role, especially among startups that do not specialize in the IoT industry. By any innovation, the campaigns and technical events promote ideas that may be forgotten. This is when startup dictionaries may come across terms such as internal financing, mass outsourcing, and collaboration.

There is the same feeling from the brand owners' point of view, and he will take the necessary profit from the competitive advantage created by affective communication to the satisfied consumer. Brand loyalty is derived from a sense of satisfaction that is the result of creating shared value. The model in **Figure 1** illustrates this relationship schematically (7). Another dimension is the interaction between brands and the customer, which causes the customer to engage with the product's contents, the function of the product in terms of professional and technical specification, and even its working model for the first time on the web. Therefore, website communicators ought to complete this connection with a proper website's proper personalization, such as flexible design. Advertisement appealing capabilities are then crucial to control this interaction. In other words, customer engagement is a set of subprocesses that use customer experience in online brand communities and value creation. It is worth noting that history helps examine the background and origins of how network marketing is created (**Figure 2**). In 1925, the theory of open systems was criticized by Alfred Luca in the study of organic systems by many researchers. Fifteen years later, the idea of an open system was improved by Ludwig Von Bartalanffy, who introduced it in management, which is widely used in entrepreneurial systems today (8). This is how Peter Singe, as the father of management science, defines learning systems in his book *The Fifth Discipline*. Advances in management science, and especially knowledge management, show that all processes can be defined as knowledge-based processes.

"The process of leveraging or customizing existing knowledge to meet the needs of users who have this personal need and seeking to create and innovate new knowledge" (9)—today, the blue ocean strategy is rooted in this concept after the emergence of cyber business only two decades after the management of comprehensive systems and chaos theory was entered into the post-modernism management theory. Researchers in cyberspace have recently observed that the collaboration space or distance working leads to even higher productivity as an emerging phenomenon called fast advertisement. When a high amount of cost is saved and the promotions are effective, they tend to use Internet cryptography to collect essential information from their consumers by the specified level of difficulty. Using online advertising technology should help this industry know users' opinions and their partners. It is worth noting that even non-consumers of an IoT encourage this competition, which creates a form of conflict between different parties and increases their attractiveness. For example, according to a published press "Binance" Exchange, well-known Bitcoins exchange mediaries experienced a security breach and were hacked on May 7, 2020. Hackers use various methods, such as phishing, to obtain large customer information such as 2FA codes and API keys, among other information (10). The hackers intend to change the price of Bitcoin, and even they were able to withdraw 7,000 bitcoins (BTC) worth more than $40 million at that time, and they controlled only 2% of the exchange resources. These frauds turn into severe conflicts and problems in the long run.

This question may also be considered in online advertising given that some companies that have branded in the field of online have also entered the field of off-line. For example, Bitcoin miners from popular website Lootbits have reached off-line

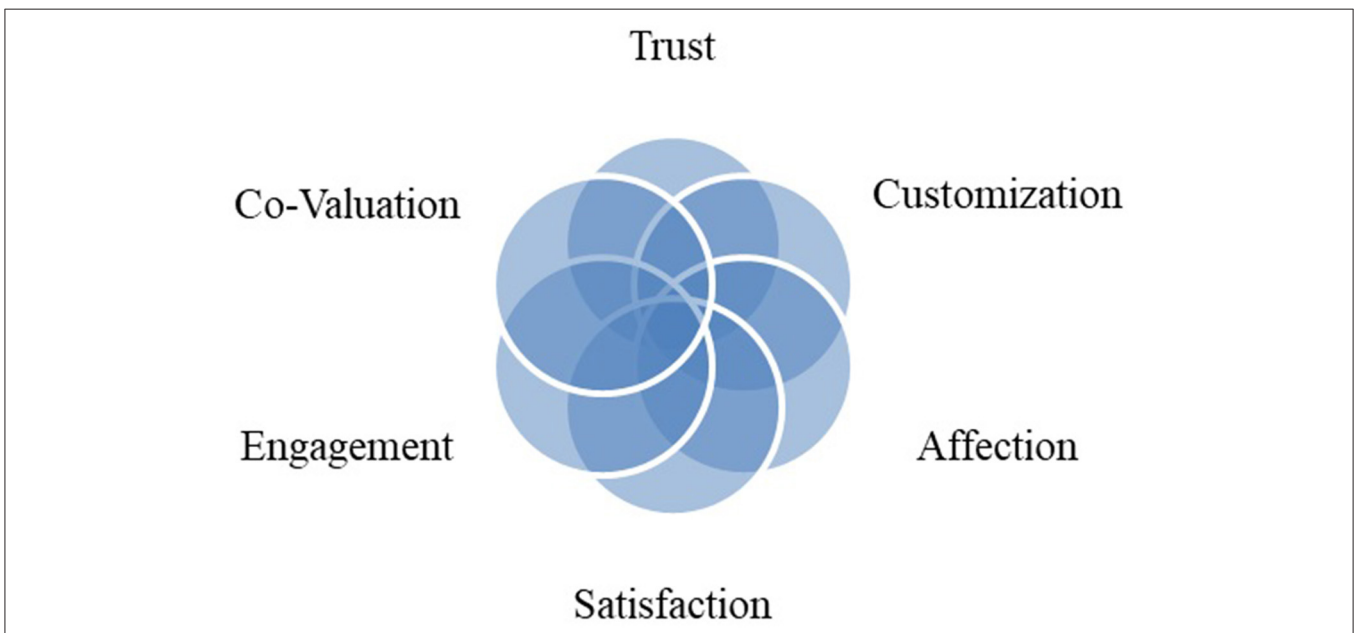


FIGURE 1 | Relationship of e-shared value, trust, and affection (7).

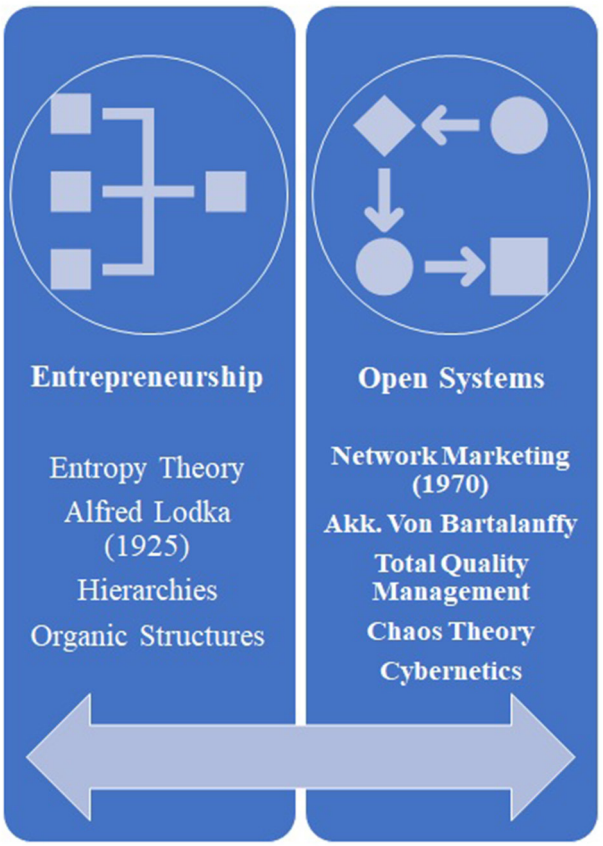


FIGURE 2 | Historical development of network marketing as the basis for the online advertisement.

activity and now operate as a gift promoter machine for soccer lotteries. Conversely, as with Paxful, there are companies that have gone online from the field of exchange business, and in the meantime, some prefer to choose a new name for their new brand, but in these cases, opponents of the existing brand believe that consumers like to register a new brand online. As a result, it subconsciously induces another brand field, shows the company's expertise, and strengthens the new name base. On the other hand, it automatically encourages the brand's stakeholders or the opponents of the new name for the brand to believe the cost of branding and the new name for the brand.

Coworking and B2B meeting groups have made significant progress in establishing virtual business offices, social networks, and telecommuting and collaboration. The argument here among startup managers is that only 1 in 10 companies succeed by developing a suitable business strategy. We should use a planner. After designing it, we can focus on a small part of the market share by differentiating or breaking the cost.

LITERATURE REVIEW

In IoT with a technological enabler, such as big data, reliability and trust have evolved a new definition. Traditionally, IT governance in the deepest layer was hierarchical. It caused lots of information silos and other inefficiency gaps. From a bottom-up point of view, citizen- and community-driven innovations are mostly needed by some stimulations and promotions. The emerging initiatives such as "Open City Dialogues" along with the government's requirements on transparency create a paradox that can best be handled by decision and policy makers (11). We believe that IoT end service users are mostly concerned about tangibility, accountability, and empathy of the service they receive. In the near future, the citizens will be experimenting with socio-regenerative developments and with emerged society rituals. Therefore, it is not surprising that incubators and sponsors of such crowd-sourced services contribute in this era, especially in the field of IoT (12).

Some of the more recent advances are using task allocation and privacy preservation algorithms. A heuristic called data aggregation error minimization (DREAM) in privacy-preserving crowdsensing uses Liyapunov stochastic programming with the random behavior of arriving sensing tasks (13, 14). It minimizes the data aggregation error of sensed tasks while preserving the privacy for a user engaging the sensing tasks.

The scientific trend of IoT has not yet finished. It covers a wide range of topics, such as neural networks, meta-heuristics' algorithms, random walks, systems dynamics, or probability and mechanical statistics. These frameworks and modeling have been recently cited through various methods and criteria in web usage, too (15–20).

For scientific research in data mining, the Yao et al. study can be mentioned in which the probability theory of websites was intended (18). In this research, it is stated that the pure usage of a website is equal to the sum of the rankings of the link pages of that website. This notation is solved by the method of Markov chain by the random walking dynamic programming algorithm. In this research, according to the experimentation of several websites with.gov extensions, the integrated methods of complexity coefficient are compared with each other. The Euclidean distances were chosen as the measurement method, and the aggregation of the ranking make it possible to accurately estimate the out-of-sample error.

In another study search engine, keywords were used for the products that were sold an e-commerce firmware (21). These rankings were then evaluated by a structural algorithm called a web dictionary, in which the frequency of that word usage was measured. For obtaining the website evaluation the retention time of users, the minimum and maximum length of the keyword, the amount of feedback, and semantics of the website were connected via a neural network applying five input layers.

In addition to using the above semantic algorithms, graph theory is also accompanied by extracting the number of webpages in the log as the nodes and the external links as the facet forming a graph. The evaluation of the website traffic activities was determined by a eugenic value vector using the chain relations of the Markov model evaluation of the website traffic activities determined by a eugenic value vector using the Markov model's chain relations. It was observed that the resulting model was convex. By calculating the eugenic factor of the site's establishment date, taking website updates and the site ranking

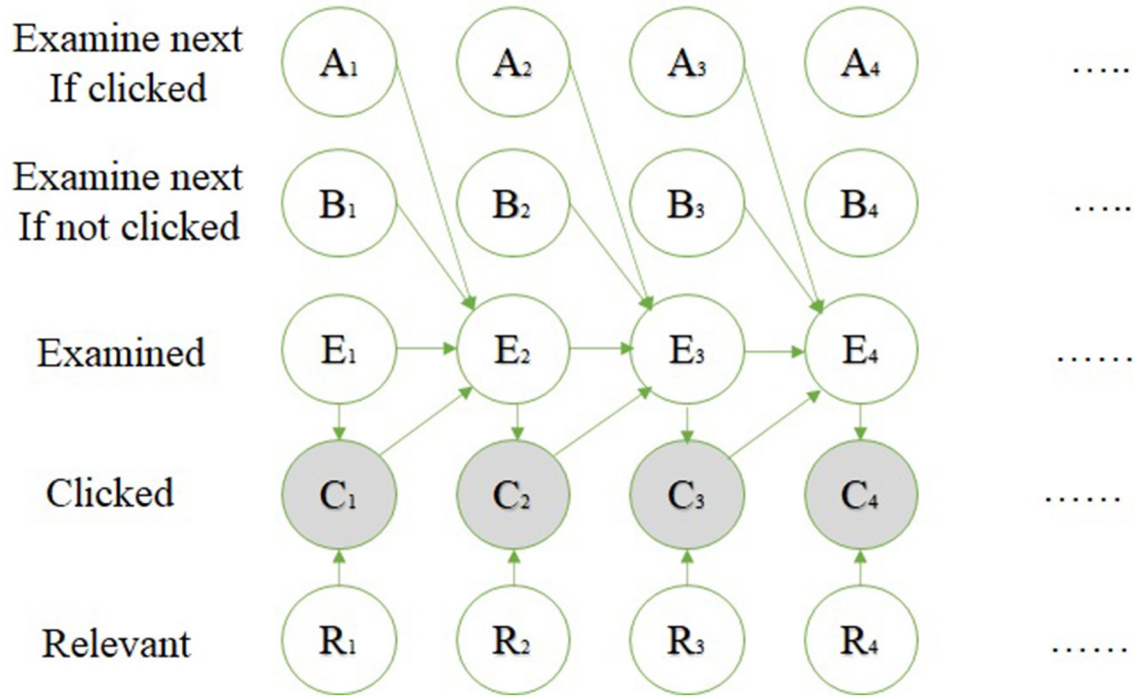


FIGURE 3 | Bayesian network (35).

relationships into account, the problem was formulated, and their time weighting in terms of added value was found.

Identifying the criteria for the success of websites through a questionnaire survey that positively increases the service level is mentioned elsewhere (22–24). Criteria, such as overall website performance, reliability, information content, and source credibility are examined separately, and the effects of other unseen characteristics were experimented. To answer the research question, structural modeling is proposed. Finally, 13 components of this study are reduced to four elements by principal component analysis (PCA). Ultimately, high-ranking websites are determined out of the 80 e-commerce websites belonging to Persian Gulf countries. In a controlled manner, these criteria are selected and finally ranked by selected individuals from selected countries with high Internet penetration rates in their countries, such as Iran, UAE, and Qatar. This data is then used to provide a theoretical framework for ranking e-commerce websites. The four elements mentioned are attractiveness, competition, engagement and retainment, indicating the amount of marketing aspect vs. the website's attractiveness at first sight.

A challenging problem in the integrated data set and the high data volume that failed in the works, as mentioned earlier, is fraud sensitivity. The data sets meet the highest standard security level by applying the most elevated standard security level with state-of-the-art machine learning to detect changes according to the algorithm's patterns. There are two methods for fraud detection: an on-the-fly schema discovery of data patterns on the client side and a solution on the data digest between the servers. Data set integration as the promising characteristics of IoT systems has enabled high security, especially in world disaster times such as the trends in virus prepandemic peak. It seems that dynamic effects are still unknown, but an agentless system, fortunately, focuses on architecture planning with monitoring an agentless system to overcome security frauds. Anyway, in intrusion cases, it is practical to perform security checks with ICMP and Ping. Both methods mentioned suffer from the low level of transparency resulting in risking usage, sharing, and processing of data sets by third parties (26, 27).

From the security point of view, this research can also explain how to detect anomalies in networks for a learning algorithm that can handle the training and testing of the data sets by an ordinary classifier (25). In this research, a binomial classification model was developed that can detect if LAN activity is in the benign or attacking category based on different attributes. The data set that is publicly available, the KDDCUP data set (28), could be used as the learning data set for testing machine learning in IoT systems. Application of different models and methods can be advantageously used for our research data set to compare the result with the advertisement sets (29). Other research models, such as the one reported in the literature, can also be helpful for the purposes of comparison accuracies (30). This data set is also discussed in the open-source data repositories by different tools and methods, and the F-score, a measure for the practical calculation of the accuracy with the consideration recall of the results, is reported to be between 0.94 and 0.96. We believe that, by selecting the appropriate preprocessing and filtering and cleaning the raw data set, it is possible to outperform this

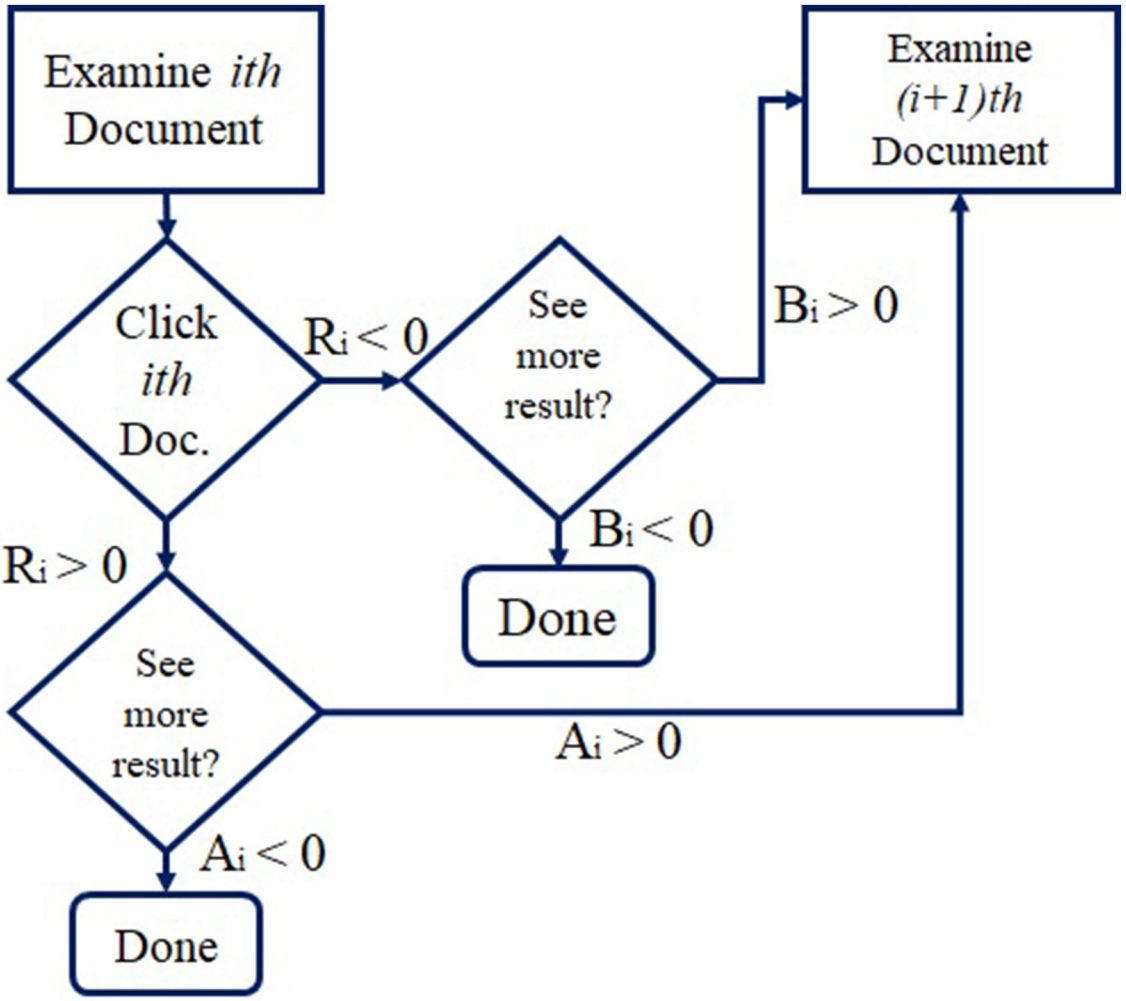


FIGURE 4 | User flow chart (35).

accuracy. It is also beneficial to hybridize the different algorithms to operationalize the same results by the least calculations and computational resources (31).

A more recent data set that was accessed from the Kaggle platform in our research is based on categorization of the attacks according to their types (32). This system's advantage relative to the binomial classes is that few overfitting risks are in the final results. The question is still unanswered as to how deep one can define each category interface to ensure the reliability of the result. Because no retrieving log file activities should be perceived as an attack, in this case, we assume that there is a difference between in- and within-group variance by the *p*-value estimates.

The other approach considered a successful one is to collect customers' intentions by mining the web server's logs to find the most linking visiting websites (17). The authors propose the use of a fuzzy web advertisement selector. Their methodology is divided into three steps: target customer clustering, fuzzy membership, and Web ad matching. For customer clustering, the authors use the neural network technique to discover backend knowledge of the web server's log files and cluster customers into different categories based on their interest. By using fuzzy rules in fuzzy inference, the authors determine the customers' categorized viewing patterns. Then, the fuzzy inference feeds the web ad selector with these patterns and enables web ad matching to suggest the corresponding advertisement to the target customer. The authors have tested their method by implementing a fuzzy web ad selector in a newspaper website. To verify the effectiveness of the fuzzy web ad selector; the authors compare click-through rate (CTR) values before and after using their method. Thus, the authors divided the newspaper web servers into two groups. One is using a web ad selector and is called the "treatment group," and the other is not using the web ad selector and is called the "control group." After running the method in newspaper websites for 2 months; there was an increase in CTR values for most sections. Businesses recognized the importance of identifying customers' needs and interests in order to provide them with right product and services. Thus, businesses are using a recommender system built to advertise products based on customer preferences. These

	The General Click Model
1	Initiate Θ let parameters satisfy $\mathcal{N}(0,1/(s+t))$
2	Construct a Bayesian inference calculator G
3	For each session S
4	M ← number of URLs in
5	Obtain the values of attribute F
6	Input Θ $\{\forall f \in F\}$ to G as prior Gaussian Dist.
7	Input the user's clicks as observations.
8	Execute to measure the posterior distributions for Θ $\{\forall f \in F\}$ and update them in Θ, End.

FIGURE 5 | GCM algorithm (35).

	Contribution Selection Algorithm
1	C: = F
2	For each $f \in F \backslash C$
3	Sample Permutations set $\{\pi_1, \ldots, \pi_t\}$ over $F \backslash C$
4	C_f: = contribution (f, C, d) = $1/t \sum \Delta_f(S_f(\pi_j))$
5	If $\max_f C_f < \Delta$
6	C: = C \ elimination ($\{C_f\}$; e , Δ)
7	go to 2
8	Else
9	Return C
10	End.

FIGURE 6 | CSA backward elimination algorithm (41).

systems are using transaction recorders, web logs, and cookies to learn customers' interests. Recommender systems are using two approaches: a content-based and a collaborative approach. A content-based approach is dependent on customer feedback. However, the collaborative approach depends on examining users' relevance and selecting a recommended partner for the target user. Thus, the collaborative approach advertises products to target users who appear in the recommendation partner only.

TABLE 1 | Literature summary.

Date	Author	Methodology
2016	Alama	Comparison of several Data Mining Algorithms for Internet of Things
2020	Lin	Real-Time Data Mining for Wireless Sensors based on IoT
2018	Yoshiaki Kawase	Queueing System for the process of Bitcoins Chain
2017	Kaur Sahi	Online Experience by Structural Equation Modeling
2016	Peng	Structural Equation Modeling
2015	Neha Derma	Data Mining and Neural Net
2015	Abdallah	Digital Marketing by Structural Equation Modeling
2018	Dijana Oreški	Pre-processing techniques for the evaluation of IoT Intrusion Detection
2020	Balaji	K-Means Feature Location for software artifacts
2017	Jiawei Zhu	Mining Information on Bitcoin Network Data
2015	Kshitij Shah	McAfee Malware detection system IBK, SVM and Adaboost
2020	J. Ni and K. Zhang	SPOON Privacy Protection and task assignment
2020	J. Xiong,	Game theory and Data Encryption
2017	T. Li	Bidding's Privacy by grouping participants
2020	L. Zhu	Collusion Resistance Mechanism in FOG computing
2020	Yang Liu, Tong Feng	Data Aggregation Error Minimization in Privacy-Preserving Crowdsensing

These two approaches have some limitations and problems, such as having no rating or feedback for new products and the rating is based on the customer point of view and style and cannot be a standard for all users.

To overcome these limitations, the authors propose a new data mining framework that is based on social networks (33). Social networks consider a rich data set that contains a lot of information can be used for advertising systems. The authors use data sets containing customer connections and those that involve customer transaction recorders. The authors use the first data set to identify customers' subgroups (each subgroup has a group of customers who know each other as friends, relatives, coworkers, etc.). The second data set is used to predict products' liking rate according to customers' past transactions. To use both sets, the authors develop a cohesion algorithm that is used to measure the number of customers sharing common interests for each subgroup based on their past transactions. This framework can target new customers by identifying to which subgroup they belong and verifying the liking rate for that subgroup. The authors have tested their proposed framework by targeting book advertising for their university library. The authors built their own social network by using email logs for their university facilities and circulation data for facilities from the library. The authors compared their approach (group-based) with other approaches: the single-based approach (assuming each user as a group); neighbor-based approach (grouping each user and his direct neighbor); and random, which randomly chose users. This study proves that their proposed framework outperforms the other approaches.

Personalizing advertising becomes a challenge for many electronic commerce companies because of the increasing number of Internet users. In addition, companies are looking to personalize advertising to reach the maximum number of customers in the world, and this cannot be done by using the traditional approaches for targeting customers by looking to their regions, age, or gender. However, better approaches to target customers by using online advertising are by analyzing the customers' preference and interest. This can be done by following the principle of "customer-based targeting" that identifies customers' interests by IP address (geographical location), navigation habits, HTTP requests, and user profiles (34).

The authors propose the advertising remote open site agent system (*AdROSA*) for automating banner advertisements. The proposed system does not violate customers' privacy as it does not store customers' details in a database. *AdROSA* analyzes the history of Internet sessions for customers along with the current session. Also, it uses web mining techniques that consist of web content and usage mining. By analyzing the history of customers' sessions and applying clustering techniques, *AdROSA* obtains aggregated sessions (called conceptual spaces) for customers with the same navigation patterns and who click on similar advertisements. *AdROSA* uses HTML content analysis of advertising to recommend the right ad matching with customers' conceptual sessions. For example, if customers belong to sport conceptual spaces, *AdROSA* analyzes advertisements that come under a sport theme and publishes the sport ad banner for customers on their current web page. The authors have tested their system by installing *AdROSA* locally on the polish tourism's website server. The system processes eight sessions and displays advertisements based on users' conceptual space. From the following table, we can see that the user's conceptual space changes based on surfing behavior (page requests). Also, *AdROSA* displays different advertising depending on users' visiting patterns and conceptual spaces.

The click model is a method that shows what the user prefers on web searching or Internet ads. This paper focuses on the previous models' problem, which uses individual queries, including the position and relevance, not a tail query. The authors mention that it is fair to build a new model that uses multiple queries or tail queries, taking in the position, local time, user agent, etc. (35–40).

The paper shows examples of previous models used in web searching and online ads, such as

- *Relevance model*: Relevance probability depends on the fact that the user clicks the URLs on the top; if the user examines the URL, then the click depends on the user's eye tracking if it is relevant.
- *Cascade model*: Depending on that the user examines all the URLs, the click depends on the order or relevance.

The authors suggest in the paper a new model called the general click model (GCM). This model has the following features:

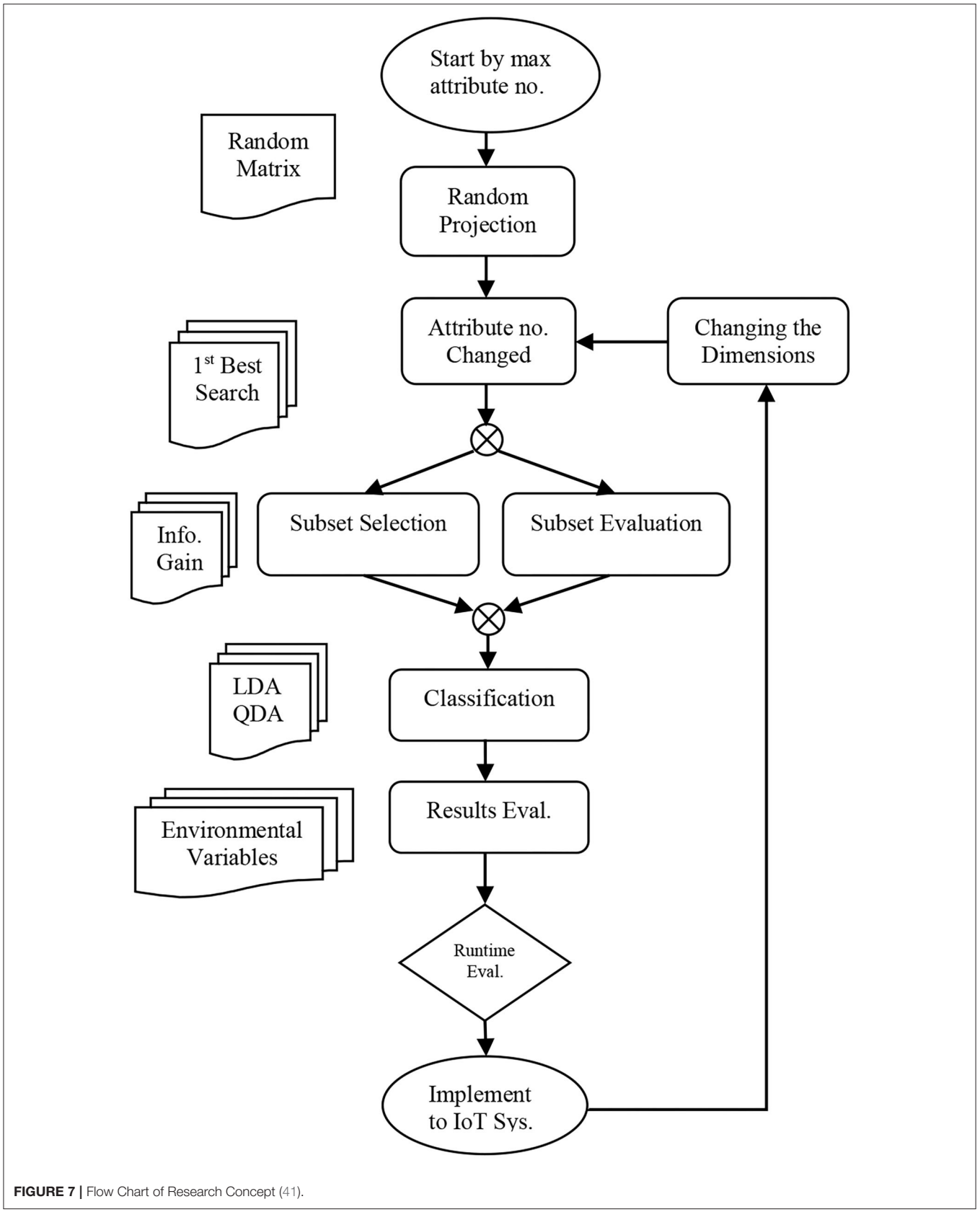


FIGURE 7 | Flow Chart of Research Concept (41).

Require: Dataset D, Attributes number(n), MSE(β)
Projection Dimensions $k= ln(n)/\beta \{d_1 d_2 \ldots d_k\}$
Subset Evaluations *s :{Cfs, Wrapper, Filter}*

1 For i:=1 to 1,2,…,s do
2 For each Subset normalize data
3 For $d' = \{d_1 d_2 \ldots d_k\}$ do
4 Do PCA and project both train and testsets into d'
5 Caculate Random Projection Matrix
6 Record the Test and Train Runtime
7 If Discrimination Experiments not significant for Normal Distribution go to 2
8 Else
9 Return Accuracy, Recall and Runtime with respect K
10 Compare with result of without Projections End.

FIGURE 8 | Algorithm pseudocode.

- Multi-bias: Using probability between many attributes.
- Across-query learning: To use one query even if it is new by learning from other queries.
- Extensibility: The ability to add or remove attributes.
- One-pass: Using an online algorithm, the output of one session is entered into the next one.
- Application (ads): GCM outperforms past works.

GCM is a general model; the previous models, such as relevance and cascade, are a special case of this model, and it is used in the nested structure, which allows this model to deal with more than the position bias. The structure of this model dealing is shown:

- The outer model: Uses the Bayesian network, shown in **Figure 3** and the user flowchart shown in **Figure 4**.
- The inner model: The query session in this model is categorized in two types:
- User-specific attributes.
- URL-specific attributes.

The algorithm of online inference using the Bayesian network and the expectation propagation method is clear in **Figure 5**. The authors performed their experiment using two general data sets: the advertisement data set using 21 attributes and the search data set. The experiments and the results applied on the ads data set are as follows:

- Evaluation on log-likelihood measurements (empirical cross entropy) used to measure one URL impression's accuracy.
- Evaluation on perplexity, which finds the accuracy for the position.
- Evaluation on R^2 by which each 1,000 URLs are blocked.

Another paper represents the contribution selection algorithm (CSA), focusing on the attributes to be selected to maximize classifiers' performance on prior unseen data as one of the feature selections benefits (41). The CSA algorithm use the filter and the wrapper approaches; it is ranked on each step based on Shapley value by a novel ranking method. The Shapley value of the player is a weighted mean over all possible subsets in the margin. The author applies the CSA using either the forward selection or the backward elimination approaches. **Figure 6** shows the algorithm. The algorithm was applied to seven data sets, including the internet advertisements data set (ads data set). Eight compared features selection schema used in the experiment are the induction algorithm with no feature selected as the baseline, SVM, different filtering, random forests, forward selection wrapper, and classification with forwarding selection

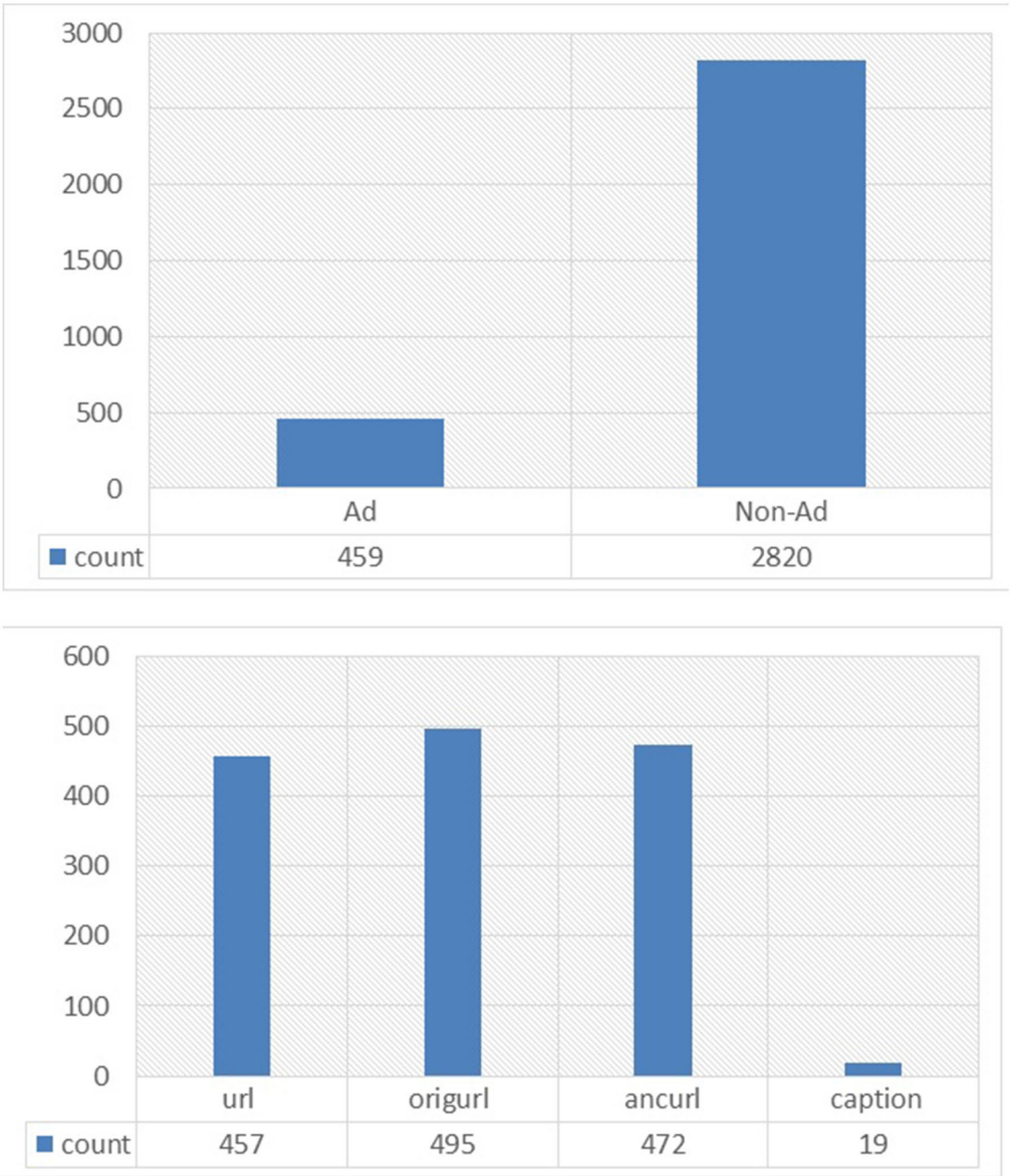


FIGURE 9 | Binomial class distribution.

or backward elimination. The results on the ads data set show that the accuracy was between 94 and 96% for all algorithms, but no feature selected by wrapper, the 1NN algorithm, was outstanding.

In another study, random projection (RP) was compared, and it is an attractive approach for dimensional reduction, compared with PCA, a popular approach (42). Five real-world data sets are used in this experiment, including the ads data set. The techniques used in the experiment to compare RP and PCA decision trees (C4.5) are nearest neighbors methods (1NN and 5NN) and linear SVM light. The authors used data sets with small sizes and others with large sizes, and they used low-dimensional data sets and other high-dimensional data sets. The ads data set was high-dimensional and large. They divided the data set into training and test subsets, depending on the size of the data set. The results show that the PCA is expensive

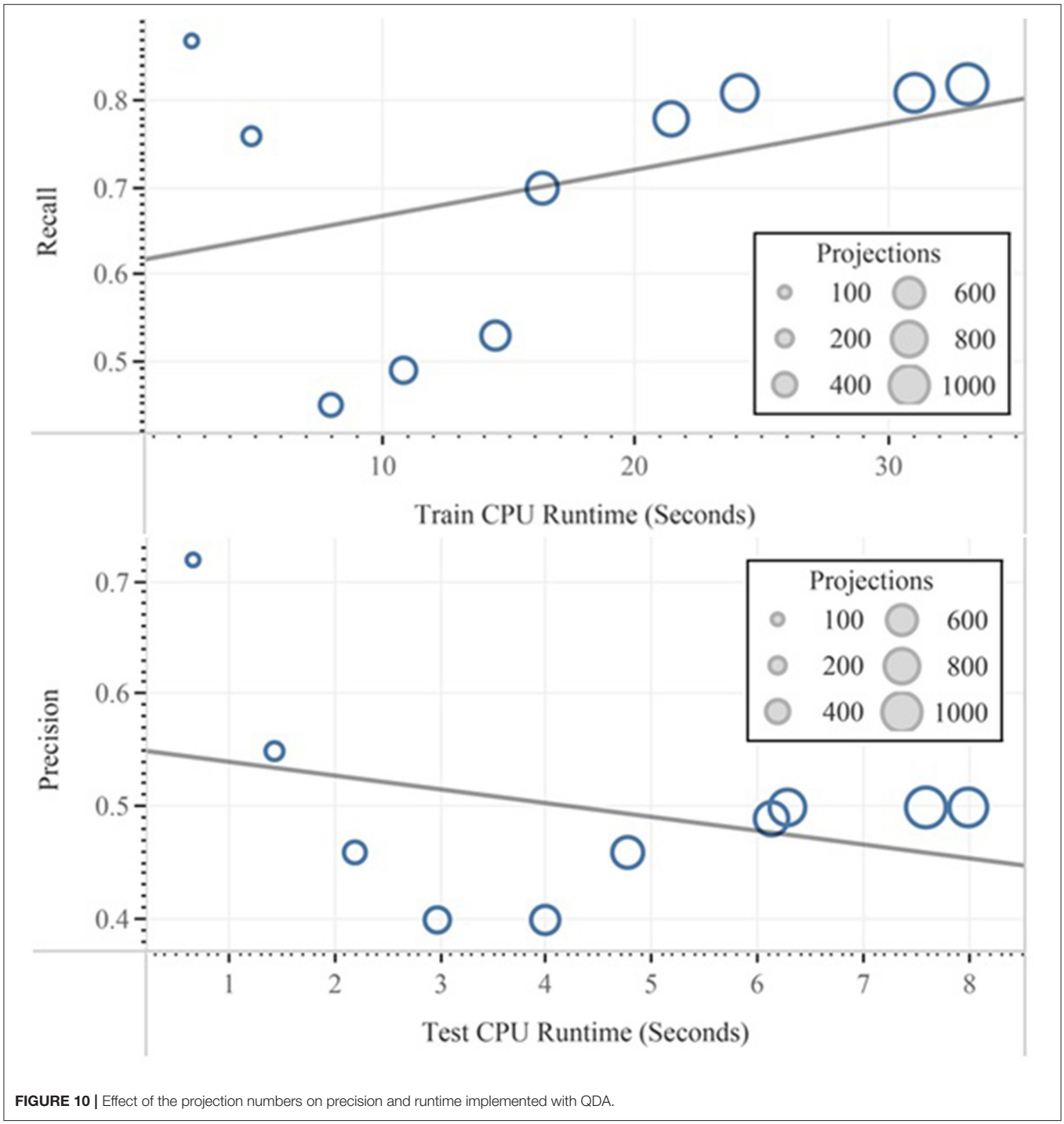


FIGURE 10 | Effect of the projection numbers on precision and runtime implemented with QDA.

computationally but gives more accurate results compared with RP, which is the cheapest computationally but has some good characteristics. Using RP improves performance by increasing the dimensions, doing the best with NN methods, and doing well with SVM. It has a problem with the decision trees (C4.5). In the **Table 1**, a summary of the research literature's methods and goals is mentioned to overview the methodologies of the literature.

RESEARCH METHOD

Discriminant analysis is widely used in engineering fields, such as electrical, vibration, and control engineering, but in this study, it is implemented in web usage and IoT (39, 43–45, 47–53). This study's objective lies in selecting a complex advertisement data set from those several hundred attributes and selecting the advertisement from a non-advertisement picture.

TABLE 2 | Classifiers accuracy for the different classification methods of each attribute selection method.

Attribute selection	Classifiers	Correctly classified instances	Incorrectly classified instances	Kappa statistic
Cfs	Decision tree (J48)	97.23%	2.78%	0.88
	k-Nearest Neighbors (IBk)	**97.76%**	2.24%	**0.91**
	Multilayer Perceptron (NNs)	96.68%	3.32%	0.86
Filter	Decision tree (J48)	97.21%	2.78%	**0.88**
	k-Nearest Neighbors (IBk)	**97.49%**	2.51%	0.90
	Multilayer Perceptron (NNs)	97.39%	2.60%	0.89
Wrapper	Decision tree (J48)	85.11%	14.89%	0
	k-Nearest Neighbors (IBk)	**85.11%**	14.89%	0
	Multilayer Perceptron (NNs)	85.11%	14.89%	0

Bold values indicates the best True Positive Rate and their kappa statistics.

TABLE 3 | Errors of classifying Ads dataset using different classifiers and attribute selection methods.

Evaluator	Classifier	MSE	RSME	Relative absolute error	Root relative squared error
Cfs	Decision tree (J48)	0.057	0.165	23.22%	46.3%
	k-Nearest Neighbors (IBk)	**0.032**	**0.149**	13.13%	**41.9%**
	Multilayer Perceptron (NNs)	0.039	0.164	16.04%	46.2%
Filter	Decision tree (J48)	0.055	0.163	22.51%	45.95%
	k-Nearest Neighbors (IBk)	0.032	0.147	13.28%	41.2%
	Multilayer Perceptron (NNs)	0.044	0.163	18.08%	45.6%
Wrapper	Decision tree (J48)	0.244	0.356	99.90%	100%
	k-Nearest Neighbors (IBk)	0.244	0.356	99.90%	100%
	Multilayer Perceptron (NNs)	0.244	0.356	100%	100%

According to **Figure 7**, the proposed algorithm method is performed in the form of classification problem with subset evaluation and a selection step. It is started by a maximum number of attributes followed by unsupervised filtering with RP, which is supposed to reduce the number of attributes. By reduction of attribute numbers in the Weka, a new subset is sorted with less dimension accompanied with new information gain. By reducing the attributes, it is expected to result in less runtime. The RP (described in **Figure 8**) filter uses the searching

TABLE 4 | Classifiers accuracy for the FLDA classification method of each attribute selection method.

Random projections	MSE	Precision	Recall	Test runtime	Train runtime
100	0.05	0.92	0.78	0.61	2.48
200	0.06	0.96	0.79	1.08	4.52
300	0.07	0.96	0.82	1.75	7.65
400	0.11	0.95	0.84	2.34	10.74
500	0.16	0.94	0.85	2.87	13.47
600	0.27	0.9	0.84	3.47	16.16
700	0.46	0.86	0.81	3.96	19.54
800	0.46	0.85	0.78	4.33	22
900	0.46	0.85	0.8	5.25	25.86
1,000	0.47	0.87	0.79	5.48	28.67

TABLE 5 | Errors of classifying ads dataset using QDA attribute selection method.

Random projection	MSE	Precision	Recall	Test runtime	Train runtime
100	0.07	0.72	0.87	0.65	2.42
200	0.12	0.55	0.76	1.42	4.78
300	0.15	0.46	0.45	2.18	7.92
400	0.17	0.4	0.49	2.96	10.79
500	0.18	0.4	0.53	3.98	14.42
600	0.16	0.46	0.7	4.76	16.27
700	0.14	0.49	0.78	6.12	21.36
800	0.14	0.5	0.81	6.27	24.07
900	0.14	0.5	0.81	7.98	30.97
1,000	0.14	0.5	0.82	7.58	33.06

algorithm based on the best first for a lower time budget. Switching the model to new data set with FLDA, the same method as mentioned in the research work of Pang et al., a Gaussian per binary class and concentrate on pooled covariance matrix instead of per multiclass vector-based covariance matrices and means computations, data mining resources are measured and reported (45).

We have additionally tried the quadratic form QLDA for the tuning and sensitivity analysis.

We want to concretely know how the companies that are part of the IoT systems have the possibility of competition to the highest degree for their business? Second, which part of computational resources of their data mining activities are the most important and why: algorithm, runtime, accuracy, data warehouses, services, etc., and finally, is there any model framework for ad selector application available to practically align data mining with the needs of all their stakeholders?

Data Description

The Internet advertisements data set (ads data set) collects data used by researchers about advertising in websites, and it can be categorized as a standard, high-dimensional, large, and imbalanced data set. It contains a set of advertisements as images, the geometry of the image, the anchor, the anchor text, and other image properties. The data set available in UCI contained 1,558 attributes and 3,279 instances. The predicted class in this data set

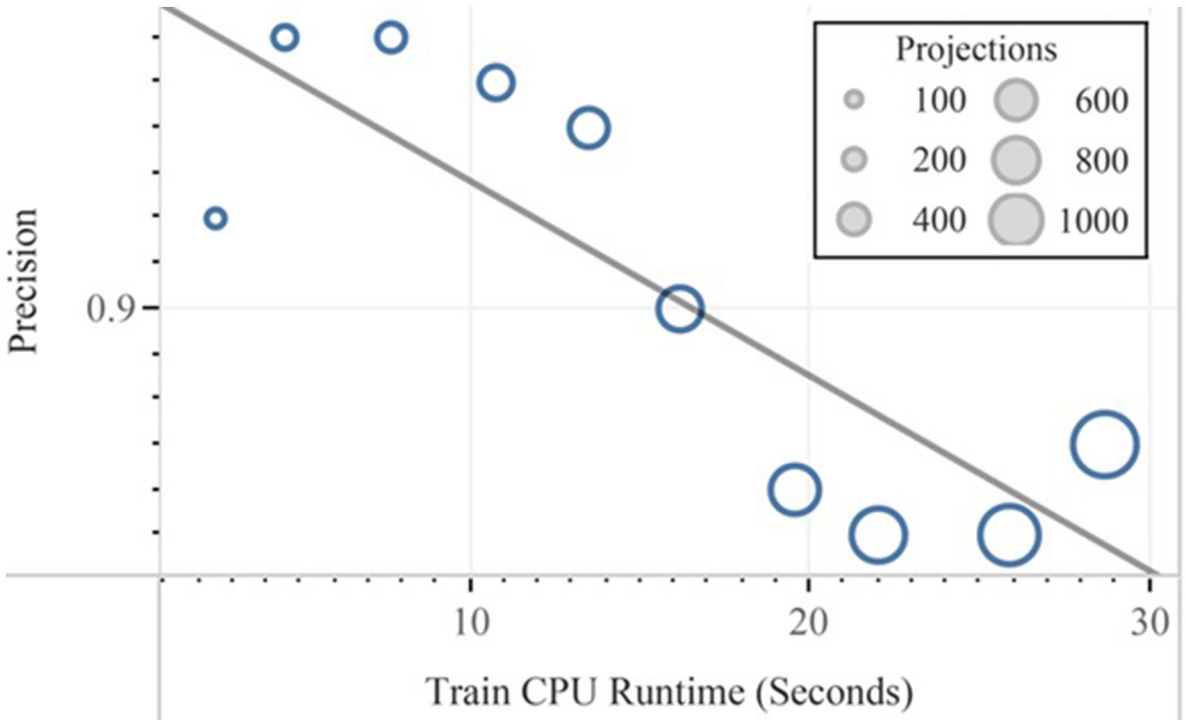


FIGURE 11 | Precision vs. training CPU runtime experimented in QDA.

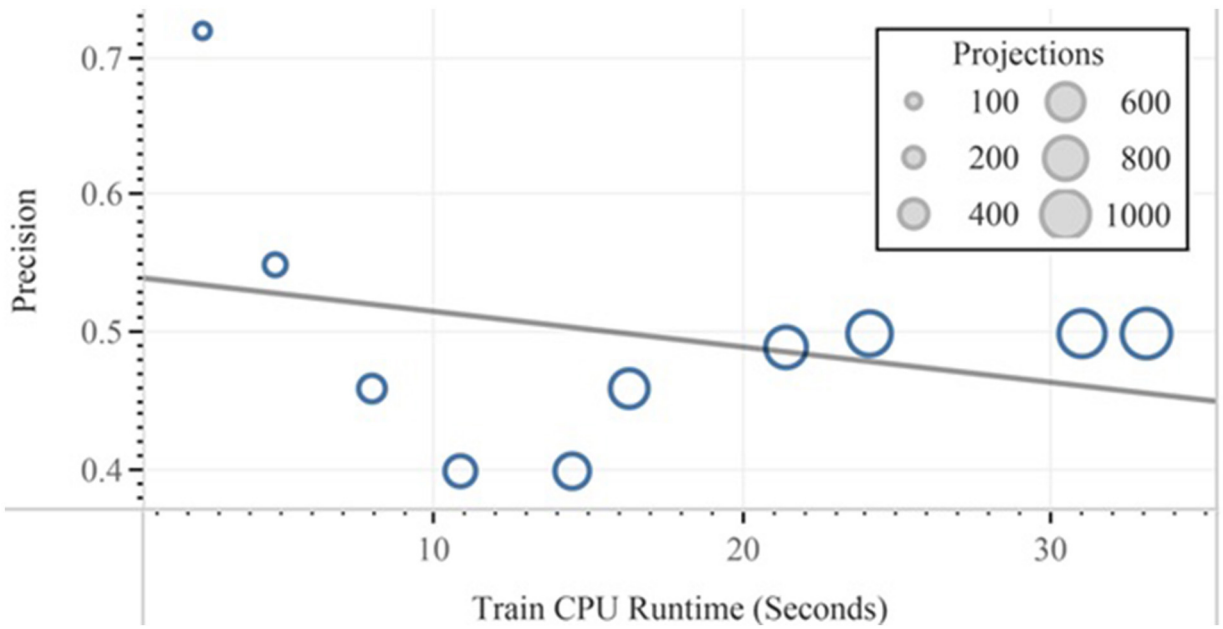


FIGURE 12 | Precision vs. training CPU runtime experimented in FLDA.

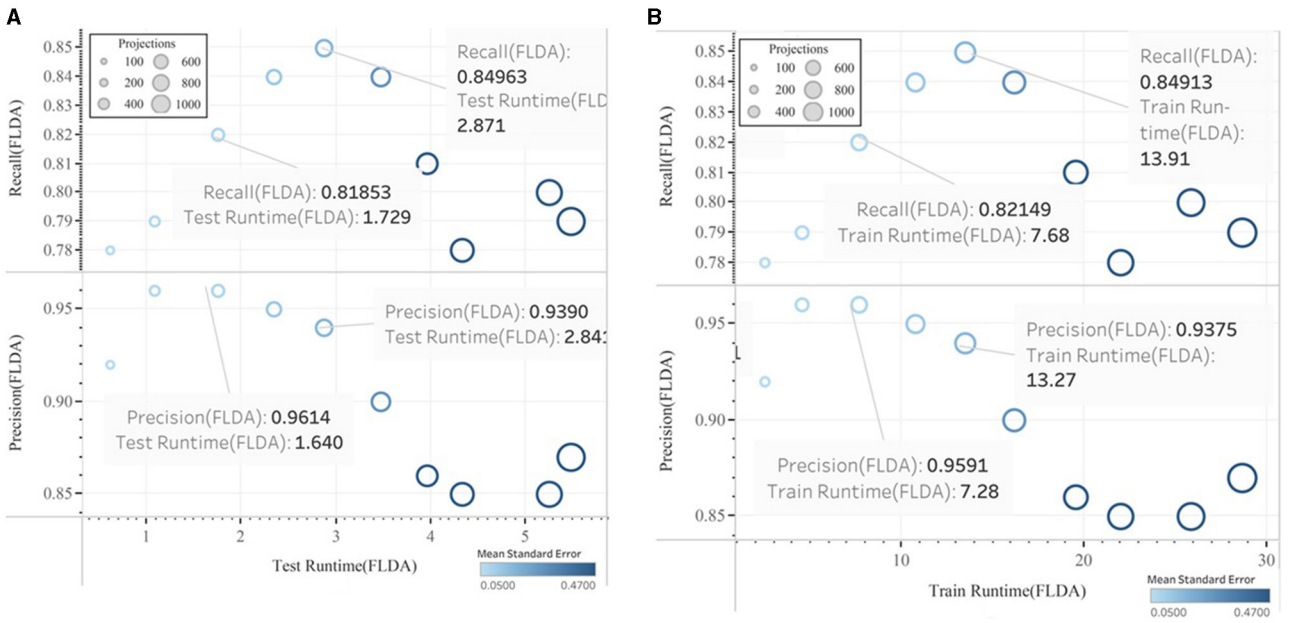


FIGURE 13 | Precision and Recall versus CPU Runtime using FLDA algorithm **(A)** Testing, **(B)** Training.

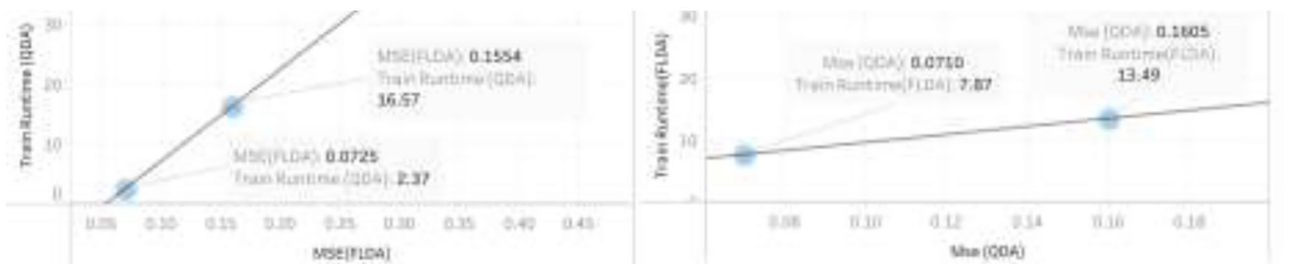


FIGURE 14 | Sensitivity analysis of mean standard error using both QDA and FLDA algorithm.

was to predict if the image is "ad" or "non-ad." The ads data set contains 2,821 non-ads and 458 ads as shown in **Figure 9**.

The ad dataset, as mentioned, contains 1,558 attributes, three of which are continuous. These are height, width, and ratio. There is 28% missing data; some of these data are from continuous attributes, which should be mentioned as "unknown." The information attributes are classified into four main categories as follows: 457 features are from URL terms, 495 features from origURL terms, 472 features from ancURL terms, and 19 features from caption terms.

Feature Selection Analysis

In feature selection, techniques are differentiated by their ability to generate other data sets from the original data from which resulted in less mean absolute error rate. In our context, this can be accompanied with the decline of the computational budget based on less runtime as well as reducing of the danger of overfitting for the IoT framework and can be later compared and approved by the above task. The only exception is that our selected data set with 2.7% error and <1% overfitting from the 10-fold cross-validation does not seem like a good candidate for the accuracy enhancement, but it can be efficiently applied for reducing the run time. This might happen on the basis of the anomaly in the distribution of the attribute's categories and also on the basis of numerical attributes. The calculation result in terms of precision and test runtime with the relevant 100 to 1,000 number of projections are represented in **Figure 10**. The data are filtered, and default settings for the calculation of the QDA were implemented.

The ads data set contains a large amount of data. The data are reduced using the attribute selection method. Feature selection or variable subset selection is the technique of selecting a subset of relevant features for classification. By removing most irrelevant features from the data, feature selection helps improve the performance of classification.

Five techniques of attribute selection were applied to the ads data set. These are Cfs, Info gain, PCA, filter, and wrapper subsets. One of the indicators of the performance of the attribute

selection is to determine the number of attributes that were selected by the given attribute selection method. In this manner, the run of the Cfs method results in selecting 24 attributes. The filter subset method gives us the minimum result to only 10 attributes. Unfortunately, the PCA method gives us a big number of attributes reaching 300 attributes. The other two methods (info gain and wrapper) failed to select attributes where it shows a full number of attributes (1,558 attributes). In conclusion, the smaller the number of attributes, the faster the results we get and the easier the model building for classification. This is in the condition of using a variety of data, which is applied in the ads data set.

The results of this work show that the combination of Cfs attribute selection method and the k-nearest neighbor method outperforms any other combination. See the below table (**Table 2**).

Classification Analysis

Classification is a data mining technique that assigns instances in a data set to target classes. The aim of classification is to accurately predict the target class for each instance in the data. Three different classification methods were used for each one of the three-attribute selections. These are the multilayer perception (NNs), the k-nearest neighbors algorithm (IBk), and the decision tree (J48). The k-nearest neighbors algorithm was set to run with $k = 1$ because there was no significant difference for higher k number. The large number of the ads data set allows us to use a Weka default value of 66% training and 34% testing set to apply the decision tree method. This is true for the other two methods of classification adopted in this work. The number of instances used for testing (34%) results in 1,115 instances out of 3,279. We started the analysis of the multilayer perception algorithm based on 1 fold in which the system divides the data automatically into 66% as a training set and 34% as a testing set. For further investigation, we redo the same technique using 10 fold cross validation (Weka default value). Using this technique, we used the whole set of instances, 10% each time as a training set. The results were significant and confirmed that the k-nearest neighbors' algorithm outperforms the other methods. Other indicators for the accuracy of classification techniques are the mean of absolute, the root mean squared, and the relative errors. All of these indicators are calculated using Weka and presented in **Table 3**. We found that the lowest mean absolute error is found in using the k-nearest neighbors (IBk) method (0.0321). The root mean squared error of it is around (0.14). This method has a root relative square error of (41.89%). An algorithm that has a low error rate is preferred as it has the most powerful classification technique.

Alternatively, the Kappa statistic (46) is used to assess the process for the classification method. It reflects the difference between the actual agreement and the agreement expected by chance; for example, a Kappa of 0.91 means there is 91% better agreement than by chance alone. Using the Kappa statistic criteria, we found that the accuracy of the three classification methods used in this study is substantial because the Kappa statistic for each one is more than 0.88 as shown in **Table 3**.

RESULTS AND DISCUSSION

As this research concern was partly on the resources relevant to the IoT, it is favorable to represent which part of computational resources of their data mining activities are impacted by the drift concept: algorithm, runtime, accuracy, data warehouses, services, etc., or is there any model framework for ad selector application available to practically align data mining with the needs of all their stakeholders. This is also important to understand the technical difference in crowdsourcing and other data sets.

In **Table 4**, the experiment results by changing the random projection on the basis of the FLDA are represented. This can be performed by the Weka experimenter module, in which the random projections are set conditionally to get the significant value for the FLDA parameters algorithm. The model performance for QDA as an evaluation method is used to assess model performance for the 10 different RP algorithms used in this work. **Table 5** represents the runtime and precision resources for the QDA algorithm. The results on these tables clearly illustrate the strength and weakness of the two methods numerically. These results are imported to the tableau software for the experimentation of these methods in regards to their runtime and accuracy sensitivity by maintaining the mean standard error in a constant Nivea.

Finally, after the feature reduction, subsequent classification with the help of linear and quadratic discrimination analyses for the class "ads" is represented. In **Figures 11**, **12**, the performance vs. runtime of FLDA and QDA are represented where their mean standard errors are equal.

The results of this work show that, in the combination of FLDA and QDA by equal mean standard error, the FLDA with shorter runtime outperforms any other combination in terms of accuracy, recall, and runtime. The runtime was decreased from 16.57 to 13.49 in addition to the decreasing of the accuracy from 96 to 94% (**Figures 13**, **14**).

CONCLUSION

This article studied the typical data mining resources such as precision, runtime and mean standard error for an IoT framework, especially healthcare applications in which reliability, responsiveness, and availability are vital. For this reason, we have applied different classification algorithms based on an Internet advertisement data set. The best algorithm for classifying is found to be FLDA with an accuracy of 94%. It has the mean absolute error at 0.16. The other classification methods gave us satisfactory results but more calculation time; however, in this work, the calculation time was decreased 20 times. These results suggest that, among the data mining techniques tested in this work, FLDA methods can significantly improve the classification methods for use in different areas, especially the IoT. This is important because, with 2% accuracy costs, the runtime can be best achieved. This is very important when there is a reported connection between ads and service reliability or, for instance, trust among empathy and anchor text. This result has been studied sizably by the feature reduction method and results were analyzed by the Fischer

linear discriminant analysis in the WEKA Experimenter module. Before recognizing the combination of the two data processing methods, the purpose of the combination should be determined, which can be used to allocate higher process time with low memory for the IoT project and drift learning. If this does not exist, we specify its constituent elements. In collective outsourcing projects, we are not opposed to technological constraints for knowledge orientation. Discriminate analysis provides a progressive framework for assessing classification techniques' performance by changing the RP attributes and other hyperparameters. This cannot only be advantageously coupled with the IoT unless the weka-based experimenter model is implemented. It is vital to include other comparison measures into the evaluation process even though using margin curves for a given data set needed more research work to determine which classification method is the best. What is its optimal decision threshold? It also required using more empirical data sets rather than using only the ads data set. This will be in the benefit of solving real-life problems in healthcare applications.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

TG and MH: research concept and methodology. AA, NN, and JA: review and editing. MH and KA: supervision. TG, MH, AO, and KA: original draft. TG, MH, and AO: software and visualization. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Universiti Kebangsaan Malaysia and the support given by the Deanship of Scientific Research, Qassim University for funding the publication.

REFERENCES

1. Yoshiaki K. A batch-service queueing system with general input and its application to analysis of mining process for bitcoin blockchain. In: *IEEE Conferences on Internet of Things*. Halifax, NS: IEEE (2018). p. 1440–7.
2. Jiawei Zhu PL. Mining information on bitcoin network data. In: *IEEE International Conference on Internet of Things*. Exeter, UK: IEEE (2017). p. 999–1003.
3. Shah K, Singh DK. A survey on data mining approaches for dynamic analysis of malwares. In: *International Conference on Green Computing and Internet of Things (ICGCloT)*. Greater Noida: IEEE (2015). p. 495–9). doi: 10.1109/ICGCIoT.2015.7380515
4. Lakshmi S, Anand N, Mathiyalaga B. An efficient scheme for secure feature location using data fusion and data mining in Internet of Things environment. *Softw Pract Exper.* (2020) 1–16. doi: 10.1002/spe.2805
5. *Semrush*, *2021*. (2016). Available online at: http://www.semrush.com (accessed January 25, 2021).
6. Kaur Sahi G. Predicting customers recommendation from Co-creation of value, customization and relational value. *J Decis Makers*. (2017) 42:19–3. doi: 10.1177/0256090916686680
7. Jevons C, Buil I, Merrilees B, de Chernatony L, Introduction: thought leadership in brand management. *J Business Res.* (2013) 66:1–3. doi: 10.1016/j.jbusres.2011.07.016
8. von Bertalanffy L. *General System Theory: Foundations, Development, Applications*. Dallas, TX: George Braziller Inc. (1969).
9. Senge PM. *The Fifth Discipline: The Art and the Learning of Organization*. New York, NY: Doubleday Dell Publishing Group (1990).
10. Wood A. *Hackers Withdraw 7,000 Bitcoins in Binance Crypto Exchange Security Breach*. Retrieved from Cointelegraph (2009). Available online at: https://cointelegraph.com/news/hackers-withdraw-7-000-bitcoins-in-binance-crypto-exchange-security-breach (accessed January 25, 2021).
11. Akhtaruzzaman M, Hasan MK, Rayhan Kabir S, Siti Norul Huda SA, Sadeq MJ, Hossain E. HSIC bottleneck based distributed deep learning model for load forecasting in smart grid with a comprehensive survey. *Evolv Technol Energy Storage Syst Energy Syst Applic.* (2020) 8:222977–3008. doi: 10.1109/ACCESS.2020.3040083
12. Flückiger Y, Seth N. SDG indicators need crowdsourcing. *Nature*. (2016) 531:448–8. doi: 10.1038/531448c
13. Liu Y, Feng T, Peng M, Guan J, Wang Y. DREAM: online control mechanisms for data aggregation error minimization in privacy-preserving crowdsensing. *IEEE Trans Depend Secure Comput [Preprint]*. (2021). doi: 10.1109/TDSC.2020.3011679
14. Ni J, Zhang K, Xia Q, Lin X, Shen XS. Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing. *IEEE Trans Mobile Comput.* (2020) 19:1317–31. doi: 10.1109/TMC.2019.2908638
15. Manaskasemsak B, Rungsawang A, Yamana H. Time-weighted web authoritative ranking. *Inform Retrieval.* (2011) 14:133–57. doi: 10.1007/s10791-010-9138-4
16. Cowart K, Goldsmith R. The influence of consumer decision-making styles on online apparel consumption by college students. *Int J Consum Stud.* (2007) 31:639–47. doi: 10.1111/j.1470-6431.2007.00615.x
17. Bae, SM, Park SC, Ha SH. Fuzzy web ad selector based on web usage mining. *IEEE Intell Syst.* (2003) 18:62–9. doi: 10.1109/MIS.2003.1249171
18. Bao Y, Feng G, Liu T-Y, Ma Z-M, Wang Y. Ranking websites: a probabilistic view. *Internet Math.* (2006) 3:295–320. doi: 10.1080/15427951.2006.10129125
19. Xiong J, Ma R, Chen L, Tian Y, Li Q, Liu X, et al. A personalized privacy protection framework for mobile crowdsensing in IIoT. *IEEE Trans Ind Informat.* (2020) 16:4231–41. doi: 10.1109/TII.2019.2948068
20. Lin L. Real time auxiliary data mining method for wireless communication mechanism optimization based on Internet of Things system. *Comput Commun.* (2020) 160:333–41. doi: 10.1016/j.comcom.2020.06.021
21. Verma N. E-commerce website ranking using semantic web mining and neural computing. In: *International Conference on Advanced Computing Technologies and Applications*. Vile Parle: Elsevier (2015). p. 45–51. doi: 10.1016/j.procs.2015.03.080
22. Engelhardt-Nowitzki C, Staberhofer, Forschungsbeiträge zur Logistik. Potentiale von softwaregestützten Produktkonfiguratoren zur Steigerung der Anpassungsfähigkeit von Unternehmen: State-of-the-Art-Erhebung aus Wissenschaft und Praxis. In: *Düren and Maastricht*. Germany: Shaker Verlag. (2010).
23. Peng L, Liao Q. Factors affecting female user information adoption: an empirical investigation on fashion shopping guide websites. *Electronic Commerce Res.* (2016) 16:145–69.
24. Abdallah S, Jaleel B. Website appeal: development of an assessment tool and evaluation framework of e-marketing. *J Theor Appl Electron Commerce Res.* (2015) 10:45–62. doi: 10.4067/S0718-18762015000300005
25. Thaseen, IS, Kumar CA. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *J King Saud Univ Comput Inform Sci.* (2017) 29:462–72. doi: 10.1016/j.jksuci.2015.12.004
26. Zhu L, Li M, Zhang Z. Secure fog-assisted crowdsensing with collusion resistance: from data reporting to data requesting. *IEEE Internet Things J.* (2019) 6:5473–84. doi: 10.1109/JIOT.2019.2902459

27. Sudhanshu Chhabra RM. Data mining on social networks for target advertisement in automobile sector. In: *International Conference on Infocom Technologies and Unmanned Systems (ICTUS'2017)*. Dubai: IEEE (2017). p. 166–70. doi: 10.1109/ICTUS.2017.8285998
28. KDD Cup 1999 Data. *Kdd.ics.uci.edu*, 2021. Available online at: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html (accessed January 25, 2021).
29. Alam F, Mehmood R, Katib I, Albeshri A. Analysis of eight data mining algorithms for smarter Internet of Things (IoT). *Proc Comput Sci.* (2016) 98:437–442. doi: 10.1016/j.procs.2016.09.068
30. Jinner. *Intrusion Detection Classification by Jinner*. Kaggle.com, 2021 (2016). Available online at: https://www.kaggle.com/what0919/intrusion-detection-classification-by-jinner (accessed January 25, 2021).
31. Oreški D, Andročec D. Hybrid data mining approaches for intrusion detection in the internet of things. In: *2018 International Conference on Smart Systems and Technologies (SST)*. Osijek: IEEE (2018). p. 221–6. doi: 10.1109/SST.2018.8564573
32. Sampadab. Network Intrusion Detection. *Kaggle.com*, 2021 (2018). Available online at: https://www.kaggle.com/sampadab17/network-intrusion-detection (accessed January 25, 2021).
33. Yang W-S. Dia J-B, Cheng H-C, Lin H-T. Mining social networks for targeted advertising. In: Sciences S, editor. *Proceedings of the 39th Annual Hawaii International Conference*, 2021. Hawaii: IEEE (2006). p. 6.
34. Kazienko P, Adamski M. AdROSA—adaptive personalization of web advertising. *Inform Sci.* (2007) 17711:2269–95. doi: 10.1016/j.ins.2007.01.002
35. Zhu ZC. A novel click model and its applications to online advertising. In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. New York, NY: ACM Conference (2010). p. 321–30. doi: 10.1145/1718487.1718528
36. Nurelmadina, N, Hasan MK, Memon I, Rashid Saeed A, Zainol Ariffin KA, Sayed Ali E, et al. A systematic review on cognitive radio in low power wide area network for industrial IoT applications. *Sustainability*. (2021) 13:338. doi: 10.3390/su13010338
37. Hasan MK, Ahmed MM, Hashim AH, Razzaque A, Islam S, Pandey B. A novel artificial intelligence based timing synchronization scheme for smart grid applications. *Wireless Pers Commun*. (2020) 114:1067–84. doi: 10.1007/s11277-020-07408-w
38. Hasan MK, Mohamud Ahmed M, Musa SS, Islam S, Sheikh Abdullah SNH, Hossain E, et al. An improved dynamic thermal current rating model for PMU-based wide area measurement framework for reliability analysis utilizing sensor cloud system. *Reliability Sensor Cloud Syst Applic*. (2021) 9:14446–58. doi: 10.1109/ACCESS.2021.3052368
39. Hasan MK, Islam S, Sulaiman R, Khan S, Hashim AH, Habib S, et al. Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access*. (2021) 9:47731–42. doi: 10.1109/ACCESS.2021.3061710
40. Memon I, Shaikh RA, Hasan MK, Hassan R, Haq AU, Zainol KA. Protect mobile travelers information in sensitive region based on fuzzy logic in IoT technology. *Security Commun Netw*. (2020) 2020:8897098. doi: 10.1155/2020/8897098
41. Cohen S, Dror G, Ruppin E. Feature selection via coalitional game theory. *Neural Comput.* (2005) 19:1939–61. doi: 10.1162/neco.2007.19.7.1939
42. Dmitriy F, Madigan D. Experiments with random projections for machine learning. In: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Washington, DC: ACM Conference (2003). p. 517–22.
43. Nikolai H. Identification and quantification of hydraulic system faults based on multivariate statistics using spectral vibration features. *Process Eng.* (2015) 1225–8. doi: 10.1016/j.proeng.2015.08.835
44. Nikolai H. Detection and compensating sensor faults in a hydraulic condition monitoring system. SENSOR 2015 and IRS 2015. In: *Saarland: AMA Conferences*. Saarbrücken: AMA Conferences (2015). p. 641–6.
45. Pang S, Ozawa S, Kasabov N. Incremental linear discriminant analysis for classification of data streams. *IEEE Trans Syst Man Cybern.* (2005) 35:905–14. doi: 10.1109/TSMCB.2005.847744
46. Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* (1960) 20:37–46.
47. Schneider T, Helwig N, Schütze A. Automatic feature extraction and selection for classification of cyclical time series data. *tm Technisches Messen*. (2017) 84:198–206. doi: 10.1515/teme-2016-0072
48. Schneiderg Tizian KS. Condition monitoring of hydraulic systems Data Set. Retrieved from UCI Center for Machine Learning and Intelligent Systems, ZeMA GmbH (2018). Available online at: http://archive.ics.uci.edu/ml/machine-learning-databases/00447/ (accessed January 25, 2021).
49. Sahu N, Verma T, Reddy GT. Diabetes classification using fuzzy logic and adaptive cuckoo search optimization techniques. *Int J Fut Revolut Comput Sci Commun Eng*. (2017) 3:252–5.
50. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *ournal of Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
51. Dhanamjayulu C, Nizhal UN, Maddikunta PK, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. In: *IET IMAGE PROCESSING*. Institution of Engineering and Technology (2021). doi: 10.1049/ipr2.12222
52. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HM, et al. Hep-Pred: Hepatitis C staging prediction using fine gaussian SVM. *Computers, Materials and Continua* 69:191–203. doi: 10.32604/cmc.2021.015436
53. Hasan MK, Shafiq M, Islam S, Pandey B, Baker El-Ebiary YA, Nafi NS, et al. Lightweight cryptographic algorithms for guessing attack protection in complex internet of things applications. *Complexity*. (2021) 2021:5540296. doi: 10.1155/2021/5540296

Conflict of Interest: AO was employed by company Tehran Urban and Suburban Railway Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Hasan, Ghazal, Alkhalifah, Abu Bakar, Omidvar, Nafi and Agbinya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 15 October 2021
doi: 10.3389/fpubh.2021.754348

A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients

Dong Wang[1,2,3†], *Jinbo Li*[1,4†], *Yali Sun*[1,2,3†], *Xianfei Ding*[1,2,3†], *Xiaojuan Zhang*[1,2,3], *Shaohua Liu*[1,2,3], *Bing Han*[1,2,3], *Haixu Wang*[1,2,3], *Xiaoguang Duan*[1,2,3] *and Tongwen Sun*[1,2,3*]

[1] General Intensive Care Unit, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Key Laboratory for Critical Care Medicine of Henan Province, Zhengzhou, China, [3] Key Laboratory for Sepsis of Zhengzhou, Zhengzhou, China, [4] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada

Background: Although numerous studies are conducted every year on how to reduce the fatality rate associated with sepsis, it is still a major challenge faced by patients, clinicians, and medical systems worldwide. Early identification and prediction of patients at risk of sepsis and adverse outcomes associated with sepsis are critical. We aimed to develop an artificial intelligence algorithm that can predict sepsis early.

Methods: This was a secondary analysis of an observational cohort study from the Intensive Care Unit of the First Affiliated Hospital of Zhengzhou University. A total of 4,449 infected patients were randomly assigned to the development and validation data set at a ratio of 4:1. After extracting electronic medical record data, a set of 55 features (variables) was calculated and passed to the random forest algorithm to predict the onset of sepsis.

Results: The pre-procedure clinical variables were used to build a prediction model from the training data set using the random forest machine learning method; a 5-fold cross-validation was used to evaluate the prediction accuracy of the model. Finally, we tested the model using the validation data set. The area obtained by the model under the receiver operating characteristic (ROC) curve (AUC) was 0.91, the sensitivity was 87%, and the specificity was 89%.

Conclusions: This newly established machine learning-based model has shown good predictive ability in Chinese sepsis patients. External validation studies are necessary to confirm the universality of our method in the population and treatment practice.

Keywords: sepsis, machine learning, prognostication, infection, ICU patients

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Abdul Rehman Javed, Air University, Pakistan
Rajesh Kaluri, VIT University, India

***Correspondence:**
Tongwen Sun
fccsuntw@zzu.edu.cn

These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *06 August 2021*
Accepted: *20 September 2021*
Published: *15 October 2021*

Citation:
Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, Han B, Wang H, Duan X and Sun T (2021) A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. Front. Public Health 9:754348. doi: 10.3389/fpubh.2021.754348

INTRODUCTION

Although numerous studies and papers on sepsis are published every year, it remains a major challenge for patients and clinicians worldwide. Between 2002 and 2012, the proportion of sepsis patients admitted to hospitals in the European ICU remained unchanged; however, the severity of the disease increased significantly (1). The standardized sepsis-related mortality rate in China in 2015 was 67 deaths per 100,000, which was equivalent to more than 1 million deaths due to sepsis (2). Despite these alarming numbers, the public seems to lack an awareness about sepsis. An adult survey found that <30% of people are aware of the severity of sepsis, which was much lower than the proportion for cardiovascular diseases, cancer, and asthma (3).

To date, the diagnosis of sepsis has largely relied on determining the presence of infection and organ dysfunction (4). In addition, screening laboratory tests are often required to confirm the diagnosis. However, laboratory testing takes time, so treatment is further delayed (5).

The early detection and prediction of patients who may develop sepsis is essential to improve the adverse consequences of sepsis. Although there are many studies on the early predictions of sepsis, such as calcitonin, C-reactive protein, white blood cells, platelets, and lactic acid (6, 7). However, disappointingly, most studies are limited in clinical prediction (8). Since sepsis is a complex clinical syndrome, it contains a wide range of multifaceted clinical and biological features; therefore, a single clinical index may not be a good reflection of the disease state (9). There is still a lack of effective biomarker combinations that can distinguish patients with sepsis from those not affected with sepsis.

Current research mainly uses data collected by bedside monitors to determine the probability of sepsis in ICU patients. Bloch et al. constructed a sepsis prediction model based on the four vital signs of mean arterial pressure, heart rate, respiratory rate, and body temperature (10). The best area under the curve (AUC) was achieved with Support Vector Machine (SVM) with radial basis function, which was 88.38%. Guillén et al. used vital sign measurements and laboratory test results to predict the likelihood of severe sepsis in patients with sepsis during ICU hospitalization (11). The study showed that the AUROC based on vital signs data was 0.84; based on vital signs and laboratory results, the AUROC was calculated to be 0.882. Calvert et al. studied the correlation between pairs and triples of vital sign measurements and the overall trend (i.e., increase, decrease, and no change) of the measurements over time to predict sepsis in the adult ICU population disease (12). Their results show that the average AUROC measurement accuracy is 0.83, but requires a larger data set, which usually requires longer processing time. Since the above studies are based on the previous definition of the Third International Consensus Definition of Sepsis (sepsis-3), our current understanding of sepsis is of limited reference value. Nemati et al. used electronic medical record data combined with high-resolution time series of heart rate and blood pressure to dynamically predict sepsis, with an area under the receiver operating characteristic (AUROC) of 0.83–0.85 (13). Although the study is based on the third international definition of sepsis (sepsis-3), its predictive power is not significantly different from previous studies.

Machine learning has been applied to multiple healthcare fields, including diabetes, cancer, cardiology, and mental health (14–17). Most of the machine learning models and tools developed in the research environment has studied the potential of prognosis, diagnosis, or clinical componentization, thus demonstrating the prospect of developing computerized decision support tools (18, 19). In general, the use of machine learning models can improve patient safety, improve the quality of care, and reduce medical costs (20).

The application of artificial intelligence in the medical field is gaining increasing recognition in the improvement of clinical practice and achievement of personalized treatment (21, 22). This study used machine learning methods to evaluate predictive clinical indicators and biomarkers related to sepsis and to establish a model that could effectively predict sepsis early, which is necessary to identify high-risk patients and may enhance the understanding and facilitate clinical management of sepsis.

Abbreviations: ICU, Intensive care unit; LDL, Low-density lipoprotein; PT, Prothrombin time; ROC, Receiver operating characteristic; WBC, White blood cells.

MATERIALS AND METHODS

Study Population

This study was a secondary analysis of a retrospective observational study conducted from 2014 to 2016 in the intensive care unit (ICU) of the First Affiliated Hospital of Zhengzhou University. The inclusion criteria were (1) infection at the time of admission to the ICU; (2) compliance with the international consensus definition of sepsis and septic shock (Sepsis-3.0); (3) age $\geq$18 years. The exclusion criteria were (1) age $<$18 years; (2) diseases without infection status such as coronary heart disease, cardiac arrest, fracture, neoplasm, cerebral infarction, and brain injury; and (3) more than three missing data. Clinical or laboratory parameters related to infection and sepsis were collected for each patient.

Statistical Analysis

The binary variables were described as counts and percentages and were evaluated using the Chi-square test or Fisher's exact test. If the continuous variables conformed to a normal distribution, they were compared using a t-test and expressed as means $\pm$ SEM. For a non-normal distribution, the Mann–Whitney U test was used. $P < 0.05$ was considered statistically significant. The ensemble model was written Python scripting language (Version 3.6.5, Python Software Foundation, Wilmington, DE, USA, https://www.python.org).

Modeling and Feature Selection

The random forest algorithm, which belongs to the category of machine learning methods and captures non-line relationships between dependent and independent variables with high flexibility and sufficient accuracy, has been successfully applied to various fields such as the estimation of the genetic effects (23), clinical deterioration prediction (24), association estimation (25), clinical outcome prediction (26), and others (27–30). In this study, we used the random forest algorithm to predict the risk of sepsis in ICU patients by analyzing laboratory/clinic data as follows: (i) lipids, (ii) liver function, (iii) hemagglutination, (iv) blood cells, (v) renal function, and (vi) electrolyte. The essential idea of the random forest algorithm is to build multiple decision trees to reduce the correlation between trees using bootstrap aggregating or bagging, which can avoid the over-fitting problem. The random forest algorithm was written in the Python scripting language (version 3.6.5, Python Software Foundation, Wilmington, DE, USA, https://www.python.org).

Generally, models with more features will achieve higher accuracy than those with fewer features. However, in clinical practice, having more features cannot always improve the performance of the model because of irrelevant or redundant features, which may mislead the models. To recognize the key features and the optimal combination of features, we performed a random forest algorithm on different subsets of the training set. In this study, we identified 55 features, which were potential candidates for sepsis prediction. Because the number of possible feature combinations was large (2^{55}), as shown in **Figure 2**, we used the Gini importance to rank the importance of all potential features (31, 32). Specifically, a high Gini importance value was a high priority for incorporation into the model. On the basis of the Gini importance value of each feature, we performed the random forest algorithm on the various feature subsets.

Validation

In this study, we used a 5-fold cross-validation to assess the prediction performance of the model because it was the most commonly used method for machine learning-based medical problem exploration (33–37). Specifically, the available training set was divided into five roughly equal-sized subsets: the training set and the validation (or internal validation) set. Four of them were applied to fit the random forest model, and the remaining one was used to estimate the accuracy of the model.

We measured the performance of the model by applying several different indices, namely (i) AUC, (ii) accuracy, (iii) precision, (iv) recall, and (v) F1-Score, which were defined as follows:

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP}$$

$$Precision = \frac{TP}{FP + TP}$$

$$Recall = \frac{TP}{FN + TP}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Here, *TP*, *FP*, *TN*, and *FN* are the number of positive samples classified as positive (true positives), the number of negative samples classified as positive (false positives), the number of positive samples classified as negative (true negatives), and the number of negative samples classified as negative (false negatives). Briefly, we used five prediction performance indices, 5-fold cross-validation for internal validation, and the testing set for external validation to estimate the performance of the model.

RESULTS

Patient Characteristics

Our database consisted of 17,005 patients admitted to the ICU. After a series of exclusions, 4,449 adult patients were included in this study, and 3,539 patients developed sepsis. The process of cohort selection is shown in **Figure 1**. A total of 55 variables, including age, sex, red blood cell count, total cholesterol, D-dimer, and other clinical or laboratory parameters related to infection and sepsis, were collected for each patient. The baseline characteristics of the included patients are shown in **Supplementary Table 1**. We then randomly divided the patients into the training and testing sets. **Supplementary Table 2** shows the basic information compared between the two sets.

Variables of Importance

Generally, the error of the model decreased with an increase in variable selection. However, increasing the number of variables was not conducive to clinical practice. In order to identify the prominent features, we used the random forest method to select variables by using various feature subsets. Therefore, the relative importance of each feature based on the fact that the features built on the tree top contributed more to the prediction of sepsis in high-risk patients is shown in **Figure 2**. It can be observed in **Figure 3** that the error value remained at a similar degree when the number of features reached 20. Therefore, we utilized a combination of 20 selected features to predict sepsis in ICU patients (shown in **Supplementary Table 3**): neutrophils%, D-dimer, neutrophils, eosinophils, lymphocytes, albumin, white blood cells (WBCs), direct bilirubin, potassium, calcium, cholinesterase, magnesium, low-density lipoprotein (LDL), prothrombin time (PT), lymphocytes, lactate dehydrogenase (LDH), basophils%, total cholesterol (TBIL), urea, and platelets (PLT).

Next, we performed a random forest classification with the same parameters (to make the comparison possible and remove the effect of the parameters) with different subsets of features to calculate the changes in AUC values. The AUC loss value changed when we set the number of features to different values (**Supplementary Table 3**).

Classification Results

As shown in **Table 1** and **Figure 4**, on average, the random forest algorithm achieved an AUC of 0.88 (±0.04), accuracy of 0.88 (±0.03), precision of 0.90 (±0.03), recall of 0.96 (±0.01), and recall of 0.93 (±0.02) in the internal validation. For the external validation, the model gave an AUC of 0.91, accuracy of 0.87, precision of 0.89, recall of 0.95, and recall of 0.92.

DISCUSSION

Early identification and treatment of sepsis is a highly complex and multifaceted challenge (38). It requires highly skilled and well-trained human experts (39). However, with the continuous emergence of AI applications in the medical field, some of these decisions will soon be replaced by machines called "intelligence" to improve clinical practice and patient outcomes (40). Most of what we call "artificial intelligence" is machine learning, which means learning from data and using this knowledge to acquire new knowledge or skills.

This study used a supervised learning method (a machine learning method) to build a predictive model, which included 20 predictors of sepsis events predicted by the random forest method. The AUC of this newly developed model was 0.91, demonstrating good discriminative power. These prediction

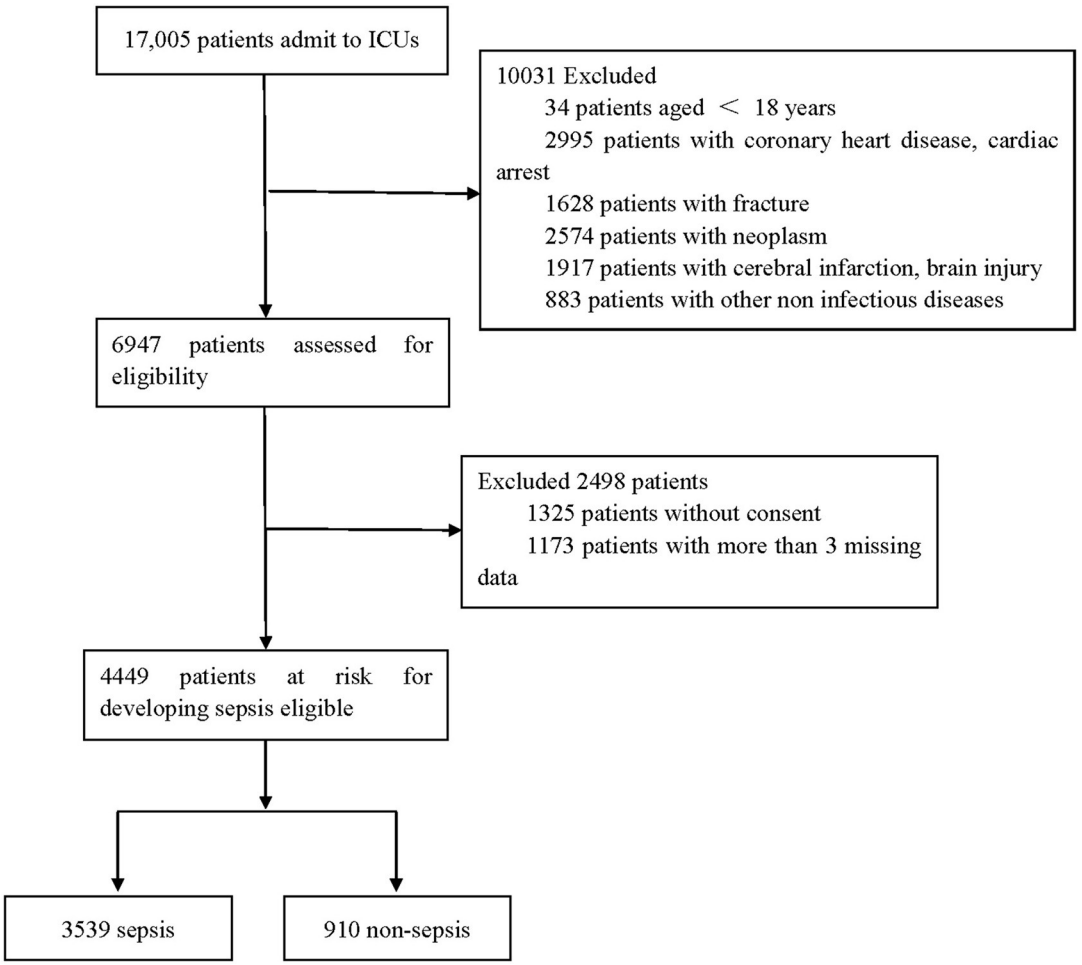


FIGURE 1 | Flow chart depicting number of patients who were included in analysis after exclusion criteria. The total included encounters were divided into those with and without sepsis.

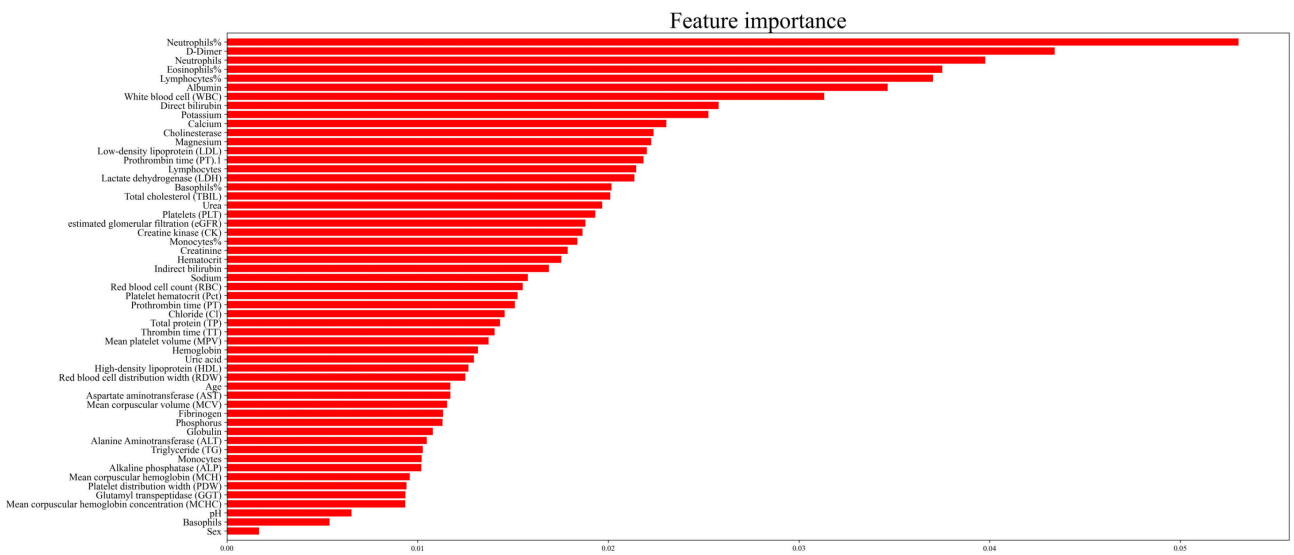


FIGURE 2 | Importance of the 20 variables included in the predictive model for sepsis events.

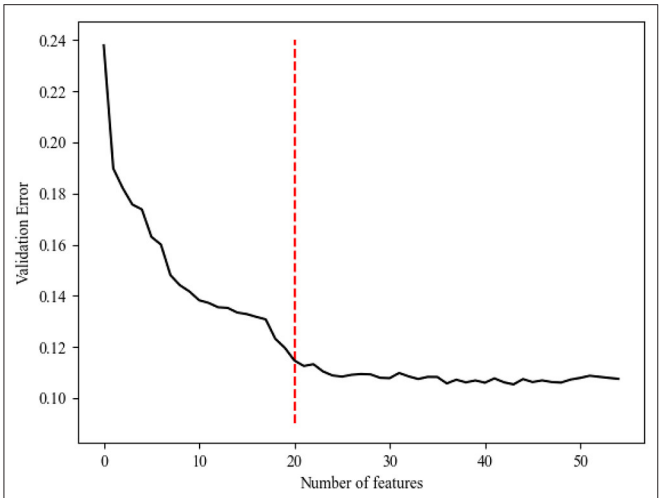


FIGURE 3 | The relationship between the cross-validation error and the number of variables.

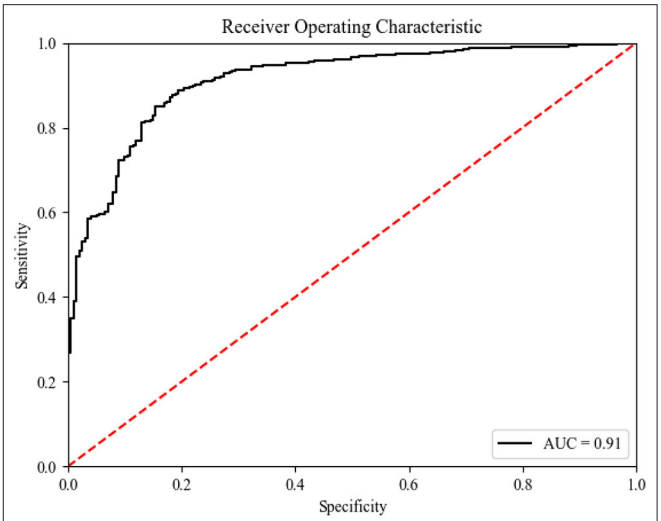


FIGURE 4 | ROC curve (of the testing set) for predicting Sepsis events using the predictive model. ROC receiver operating characteristic.

TABLE 1 | Internal and external validation results of the prediction model.

	AUC	Accuracy	Precision
Internal validation/validation set (95% CI)	0.88 (±0.04)	0.88 (±0.03)	0.90 (±0.03)
External validation/testing set	0.91	0.87	0.89
	Recall	**F1-Score**	
Internal validation/validation set (95% CI)	0.96 (±0.01)	0.93 (±0.02)	
External validation/testing set	0.95	0.92	

CI, confidence intervals; AUC, indicates area under the receiver-operating curves.

results suggest that the ensemble model with 20 key features is feasible and practical.

To our knowledge, most previous studies have developed models to predict the prognosis of sepsis. However, only few researchers have paid attention to the differences in the incidence of sepsis after infection, although it is important for clinical preventive intervention. Thomas et al. developed machine learning models for the early identification of sepsis risk (41); however, they did not obtain precise biomarkers that could be applied to clinicians. All calculations are trivial for a computer, which may limit generalization of the results to other hospitals and hospital systems. Other artificial intelligence systems such as random forest models may be a valuable tool to predict sepsis (8).

The variables in our model were mainly blood cells, lipids, liver function, hemagglutination, renal function, electrolyte, enzyme, and others. Interestingly, blood-related variables accounted for a large part of our model; the first five variables in **Figure 2** are related to the blood system. Neutrophils were an ideal choice for eliminating pathogenic bacteria because they store a large number of proteolytic enzymes that can rapidly produce reactive oxygen species to degrade internal pathogens. Hence, patients with sepsis often have neutrophil infiltration, and the degree of infiltration is related to tissue damage (42, 43). Other blood cells, including eosinophils, basophils, lymphocytes, and WBCs, are also associated with the body's defense against infection. For example, some studies have speculated that individuals with basophilic granulocytopenia have a weak resistance to infection and thus are more likely to develop sepsis (44). In addition, studies have shown that eosinophilia was a moderate marker for distinguishing SIRS from infection in critically ill patients newly admitted to the hospital, which suggested that eosinophilia may be a useful clinical tool for the prediction of sepsis (45). In addition, lymphocyte apoptosis has been recognized as an important step in the pathogenesis of experimental sepsis by inducing a state of "immune paralysis" that renders the host vulnerable to invading pathogens (46).

In the past decade, there has been a growing awareness about the role of the coagulation and fibrinolysis systems in the development of inflammation. Patients with sepsis may have common host reactions, such as coagulation, inflammation, and endothelial injury. Abnormal inflammatory and coagulation biomarkers were found to be associated with disease severity and mortality in patients with severe sepsis (47). Platelets are the main effector cells involved in blood coagulation and can promote the development of excessive inflammation, DIC, and microthrombosis (48). PT can reflect the coagulation function of the body, and D-dimer levels increase under hypercoagulable state (49). Therefore, changes in these substances may predict the occurrence of sepsis.

Sepsis is often associated with multiple organ dysfunction such as that involving the heart, liver, and kidney (50). Therefore, some indicators reflecting organ function may be used to predict the occurrence of sepsis. Albumin which is the most important protein in human plasma, maintains nutrition and osmotic

pressure. When liver synthesis is dysfunctional, its level usually decreases. Lactate dehydrogenase and urea are associated with cardiac and renal function, respectively. Patel et al. revealed an association between serum bilirubin levels and mortality during sepsis, suggesting that serum bilirubin may be a potential predictor of sepsis occurrence and death (51).

Previous studies have shown that lipids are also involved in the occurrence and development of sepsis. Yamano et al. found that low total cholesterol and high total bilirubin levels are associated with prognosis in patients with prolonged sepsis (52). Hofer et al. found that pharmacologic inhibition of cholinesterase improves survival in experimental sepsis, probably by activating the cholinergic anti-inflammatory pathway (53). The results of Feng's study suggest that a decrease in LDL-C levels is significantly associated with an increased risk of sepsis in infected patients, although the association was due to the presence of complications (54).

Although the association between electrolytes other than calcium and sepsis appears to be poorly studied, this study found that the decrease of potassium and magnesium is closely related to the occurrence of sepsis. We know that the critical illness itself is associated with a decrease in serum total calcium and free calcium levels, which is related to the severity of underlying diseases as measured by the APACHE II score. In addition, studies have shown that total and ionized hypocalcemia is more significantly associated with increased severity of infection, which suggested the role of calcium in predicting the risk of sepsis in patients with infection (55). Regarding magnesium and potassium, a study pointed out that ATP-$MgCl_2$ may be beneficial in sepsis (56). An increasing amount of evidence has suggested that potassium channels are involved in cardiovascular dysfunction in sepsis after systemic inflammation, cardiovascular dysfunction, and organ damage, and that potassium channels may affect the emergence of sepsis after infection (57). In conclusion, we believe that because sepsis is not a simple disease that can be predicted by a single marker, the biomarkers included in our model can be combined to predict the risk of sepsis in infected patients.

Our study has several limitations. First, this was a retrospective study, which had its own shortcomings, such as information bias. Second, the prediction model may have lacked generality because the 55 variables are still too few, and many other variables were omitted due to the loss of too many values. Generally, the more the variables included, the higher the prediction accuracy. Therefore, we hope to include more patients and variables in future prospective studies.

A model with 20 key features was successfully established to predict sepsis events in Chinese patients. This model has excellent ability to predict sepsis events in Chinese patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DW, YS, and JL participated in the research design and coordination and helped to draft the manuscript. XDi and XZ contributed the acquisition of data. JL, SL, BH, HW, XDu, and TS performed the data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants from the United Fund of National Natural Science Foundation of China (U2004110); Leading Talents Fund in Science and Technology Innovation in Henan Province (194200510017); Science and Technology people-benefit project of Zhengzhou (2019KJHM0001). The special fund for young and middle-aged medical research of China International Medical Exchange Foundation (Z-2018-35); The integrated thinking research foundation of the China foundation for International Medical Exchange (Z-2016-23-2001); The study of mechanism of Gabexate Mesilate in the treatment of sepsis and septic shock (2019-hx-45).

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.cn) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.754348/full#supplementary-material

REFERENCES

1. Vincent JL, Lefrant JY, Kotfis K, Nanchal R, Martin-Loeches I, Wittebole X, et al. Comparison of European ICU patients in 2012 (ICON) versus 2002 (SOAP). *Intensive Care Med*. (2018) 44:337–44. doi: 10.1007/s00134-017-5043-2
2. Weng L, Zeng XY, Yin P, Wang LJ, Wang CY, Jiang W, et al. Sepsis-related mortality in China: a descriptive analysis. *Intensive Care Med*. (2018) 44:1071–80. doi: 10.1007/s00134-018-5203-z
3. Kerrigan SW, Martin-Loeches I. Public awareness of sepsis is still poor: we need to do more. *Intensive Care Med*. (2018) 44:1771–3. doi: 10.1007/s00134-018-5307-5
4. Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med*. (2013) 369:840–51. doi: 10.1056/NEJMra1208623
5. Herzum I, Renz H. Inflammatory markers in SIRS, sepsis and septic shock. *Curr Med Chem*. (2008) 15:581–7. doi: 10.2174/092986708783769704
6. Küster H, Weiss M, Willeitner AE, Detlefsen S, Jeremias I, Zbojan J, et al. Interleukin-1 receptor antagonist and interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation. *Lancet*. (1998) 352:1271–7. doi: 10.1016/S0140-6736(98)08148-3
7. Brunkhorst FM, Wegscheider K, Forycki ZF, Brunkhorst R. Procalcitonin for early diagnosis and differentiation of SIRS, sepsis, severe sepsis, and septic shock. *Intensive Care Med*. (2000) 26(Suppl. 2):S148–52. doi: 10.1007/s001340051134

8. Møller MH, Alhazzani W, Shankar-Hari M. Focus on sepsis. *Intensive Care Med*. (2019) 45:1459–61. doi: 10.1007/s00134-019-05680-4
9. Hernandez G, Bellomo R, Bakker J. The ten pitfalls of lactate clearance in sepsis. *Intensive Care Med*. (2019) 45:82–5. doi: 10.1007/s00134-018-5213-x
10. Bloch E, Rotem T, Cohen J, Singer P, Aperstein Y. Machine learning models for analysis of vital signs dynamics: a case for sepsis onset prediction. *J Healthc Eng*. (2019) 2019:5930379. doi: 10.1155/2019/5930379
11. Guillén J, Liu J, Furr M, Wang T, Strong S, Moore C, et al. Predictive models for severe sepsis in adult ICU patients. In: *2015 Systems and Information Engineering Design Symposium*. Charlottesville, VA: IEEE (2015). p. 182–7. doi: 10.1109/SIEDS.2015.7116970
12. Calvert JS, Price DA, Chettipally UK, Barton CW, Barton CW, Feldman MD, et al. A computational approach to early sepsis detection. *Comput Biol Med*. (2016) 74:69–73. doi: 10.1016/j.compbiomed.2016.05.003
13. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med*. (2018) 46:547–53. doi: 10.1097/CCM.0000000000002936
14. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci*. (2021) 7:e390. doi: 10.7717/peerj-cs.390
15. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmanna K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Hum Comput*. (2021) 1–12. doi: 10.1007/s12652-021-03154-4
16. Javed AR, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. PP-SPA: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Process Lett*. (2021) 1–18. doi: 10.1007/s11063-020-10414-5
17. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfk H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. *Hum Cent Comput Inf. Sci*. (2020) 10:40. doi: 10.1186/s13673-020-00245-7
18. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, et al. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front Neurosci*. (2015) 9:307. doi: 10.3389/fnins.2015.00307
19. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kalur R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell*. (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
20. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med*. (2019) 25:433–8. doi: 10.1038/s41591-018-0335-9
21. Dhanamjayulu C, Nizhal UN, Maddikunta PR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) 1–12. doi: 10.1049/ipr2.12222
22. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
23. Stephan J, Stegle O, Beyer A. A random forest approach to capture genetic effects in the presence of population structure. *Nat Commun*. (2015) 6:7432. doi: 10.1038/ncomms8432
24. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*. (2016) 44:368–74. doi: 10.1097/CCM.0000000000001571
25. Pannaraj PS, Li F, Cerini C, Bender JM, Yang S, Rollie A, et al. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. *JAMA Pediatr*. (2017) 171:647–54. doi: 10.1001/jamapediatrics.2017.0378
26. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA Jr, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care*. (2019) 23:64. doi: 10.1186/s13054-019-2351-7
27. Kaup AR, Nettiksimmons J, Harris TB, Sink KM, Satterfield S, Metti AL, et al. Cognitive resilience to apolipoprotein E ε4: contributing factors in black and white older adults. *JAMA Neurol*. (2015) 72:340–8. doi: 10.1001/jamaneurol.2014.3978
28. Lanspa MJ, Gutsche AR, Wilson EL, Olsen TD, Hirshberg EL, Knox DB, et al. Application of a simplified definition of diastolic function in severe sepsis and septic shock. *Crit Care*. (2016) 20:243. doi: 10.1186/s13054-016-1421-3
29. Flechet M, Güiza F, Schetz M, Wouters P, Vanhorebeek I, Derese I, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive Care Med*. (2017) 43:764–73. doi: 10.1007/s00134-017-4678-3
30. Roimi M, Neuberger A, Shrot A, Paul M, Geffen Y, Bar-Lavie Y. Early diagnosis of bloodstream infections in the intensive care unit using machine-learning algorithms. *Intensive Care Med*. (2020) 46:454–62. doi: 10.1007/s00134-019-05876-8
31. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics*. (2009) 10:213. doi: 10.1186/1471-2105-10-213
32. Li Y, Middaugh CR, Fang J. A novel scoring function for discriminating hyperthermophilic and mesophilic proteins with application to predicting relative thermostability of protein mutants. *BMC Bioinformatics*. (2010) 11:62. doi: 10.1186/1471-2105-11-62
33. Singh J, Hanson J, Paliwal K, Zhou Y. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun*. (2019) 10:5407. doi: 10.1038/s41467-019-13395-9
34. Maverakis E, Ma C, Shinkai K, Fiorentino D, Callen JP, Wollina U, et al. Diagnostic criteria of ulcerative pyoderma gangrenosum: a delphi consensus of international experts. *JAMA Dermatol*. (2018) 154:461–66. doi: 10.1001/jamadermatol.2017.5980
35. Callaham M, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA*. (2002) 287:2847–50. doi: 10.1001/jama.287.21.2847
36. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. (2018) 378:1396–407. doi: 10.1056/NEJMoa1801445
37. Cheng J, Han Z, Mehra R, Shao W, Cheng M, Feng Q, et al. Computational analysis of pathological images enables a better diagnosis of TFE3 Xp11.2 translocation renal cell carcinoma. *Nat Commun*. (2020) 11:1778. doi: 10.1038/s41467-020-15671-5
38. Helms J, Perner A. Focus on sepsis. *Intensive Care Med*. (2020) 46:1457–9. doi: 10.1007/s00134-020-06038-x
39. Komorowski M. Clinical management of sepsis can be improved by artificial intelligence: yes. *Intensive Care Med*. (2020) 46:375–7. doi: 10.1007/s00134-019-05898-2
40. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. (2019) 25:44–56. doi: 10.1038/s41591-018-0300-7
41. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. *JMIR Med Inform*. (2016) 4:e28. doi: 10.2196/medinform.5909
42. Edward SW. *Biochemistry and Physiology of the Neutrophil*. Cambridge: Cambridge University Press (1994).
43. Brown KA, Brain SD, Pearson JD, Edgeworth JD, Lewis SM, Treacher DF. Neutrophils in development of multiple organ failure in sepsis. *Lancet*. (2006) 368:157–69. doi: 10.1016/S0140-6736(06)69005-3
44. Piliponsky AM, Shubin NJ, Lahiri AK, Truong P, Clauson M, Niino K, et al. Basophil-derived tumor necrosis factor can enhance survival in a sepsis model in mice. *Nat Immunol*. (2019) 20:129–40. doi: 10.1038/s41590-018-0288-7
45. Abidi K, Khoudri I, Belayachi J, Madani N, Zekraoui A, Zeggwagh AA, et al. Eosinopenia is a reliable marker of sepsis on admission to medical intensive care units. *Crit Care*. (2008) 12:R59. doi: 10.1186/cc6883
46. Hotchkiss RS, Tinsley KW, Swanson PE, Schmieg RE Jr, Hui JJ, Chang KC, et al. Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans. *J Immunol*. (2001) 166:6952–63. doi: 10.4049/jimmunol.166.11.6952
47. Kinasewitz GT, Yan SB, Basson B, Comp P, Russell JA, Cariou A, et al. Universal changes in biomarkers of coagulation and inflammation occur in patients with severe sepsis, regardless of causative micro-organism [ISRCTN74215569]. *Crit Care*. (2004) 8:R82–90. doi: 10.1186/cc2459

48. de Stoppelaar SF, van 't Veer C, van der Poll T. The role of platelets in sepsis. *Thromb Haemost*. (2014) 112:666–77. doi: 10.1160/TH14-02-0126
49. Rodelo JR, De la Rosa G, Valencia ML, Ospina S, Arango CM, Gómez CI, et al. D-dimer is a significant prognostic factor in patients with suspected infection and sepsis. *Am J Emerg Med*. (2012) 30:1991–9. doi: 10.1016/j.ajem.2012.04.033
50. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287
51. Patel JJ, Taneja A, Niccum D, Kumar G, Jacobs E, Nanchal R. The association of serum bilirubin levels on the outcomes of severe sepsis. *J Intensive Care Med*. (2015) 30:23–9. doi: 10.1177/0885066613488739
52. Yamano S, Shimizu K, Ogura H, Hirose T, Hamasaki T, Shimazu T, et al. Low total cholesterol and high total bilirubin are associated with prognosis in patients with prolonged sepsis. *J Crit Care*. (2016) 31:36–40. doi: 10.1016/j.jcrc.2015.09.033
53. Hofer S, Eisenbach C, Lukic IK, Schneider L, Bode K, Brueckmann M, et al. Pharmacologic cholinesterase inhibition improves survival in experimental sepsis. *Crit Care Med*. (2008) 36:404–8. doi: 10.1097/01.CCM.0B013E31816208B3
54. Feng Q, Wei WQ, Chaugai S, Leon BGC, Mosley JD, Leon DAC, et al. Association between low-density lipoprotein cholesterol levels and risk for sepsis among patients admitted to the hospital with infection. *JAMA Netw Open*. (2019) 2:e187223. doi: 10.1001/jamanetworkopen.2018.7223
55. Müller B, Becker KL, Kränzlin M, Schächinger H, Huber PR, Nylèn ES, et al. Disordered calcium homeostasis of sepsis: association with calcitonin precursors. *Eur J Clin Invest*. (2000) 30:823–31. doi: 10.1046/j.1365-2362.2000.00714.x
56. Harkema JM, Chaudry IH. Magnesium-adenosine triphosphate in the treatment of shock, ischemia, and sepsis. *Crit Care Med*. (1992) 20:263–75. doi: 10.1097/00003246-199202000-00015
57. Sordi R, Fernandes D, Heckert BT, Assreuy J. Early potassium channel blockade improves sepsis-induced organ damage and cardiovascular dysfunction. *Br J Pharmacol*. (2011) 163:1289–301. doi: 10.1111/j.1476-5381.2011.01324.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Li, Sun, Ding, Zhang, Liu, Han, Wang, Duan and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 22 October 2021
doi: 10.3389/fpubh.2021.762303

Smart Cardiac Framework for an Early Detection of Cardiac Arrest Condition and Risk

Apeksha Shah[1], Swati Ahirrao[1], Sharnil Pandya[1], Ketan Kotecha[2] and Suresh Rathod[1]*

[1] Computer Science Department, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India, [2] Symbiosis Centre for Applied Artificial Intelligence, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India

Cardiovascular disease (CVD) is considered to be one of the most epidemic diseases in the world today. Predicting CVDs, such as cardiac arrest, is a difficult task in the area of healthcare. The healthcare industry has a vast collection of datasets for analysis and prediction purposes. Somehow, the predictions made on these publicly available datasets may be erroneous. To make the prediction accurate, real-time data need to be collected. This study collected real-time data using sensors and stored it on a cloud computing platform, such as Google Firebase. The acquired data is then classified using six machine-learning algorithms: Artificial Neural Network (ANN), Random Forest Classifier (RFC), Gradient Boost Extreme Gradient Boosting (XGBoost) classifier, Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree (DT). Furthermore, we have presented two novel gender-based risk classification and age-wise risk classification approach in the undertaken study. The presented approaches have used Kaplan-Meier and Cox regression survival analysis methodologies for risk detection and classification. The presented approaches also assist health experts in identifying the risk probability risk and the 10-year risk score prediction. The proposed system is an economical alternative to the existing system due to its low cost. The outcome obtained shows an enhanced level of performance with an overall accuracy of 98% using DT on our collected dataset for cardiac risk prediction. We also introduced two risk classification models for gender- and age-wise people to detect their survival probability. The outcome of the proposed model shows accurate probability in both classes.

Keywords: artificial intelligence, cardiac arrest prediction, machine learning, predictive analysis, risk classification, heart failure

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Chintan Modi, Independent Researcher, Sarnia, Canada
Praveen Kumar, VIT University, India
Abdul Rehman Javed, Air University, Pakistan

****Correspondence:***
Sharnil Pandya
sharnil.pandya@sitpune.edu.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *21 August 2021*
Accepted: *20 September 2021*
Published: *22 October 2021*

Citation:
Shah A, Ahirrao S, Pandya S, Kotecha K and Rathod S (2021) Smart Cardiac Framework for an Early Detection of Cardiac Arrest Condition and Risk. Front. Public Health 9:762303. doi: 10.3389/fpubh.2021.762303

INTRODUCTION

At present times, cardiovascular disease (CVD) is one of the most contagious illnesses. According to WHO, ∼60% of cardiac patients are Indians to suffer from CVDs. Cardiac symptoms are generally associated with dynamic changes in an individual. The vital changes in the human body may lead to cardiac problems due to misdiagnoses or improper treatment. Another sign of cardiac disease is environmental changes and lifestyles of people (1). In healthcare, CVDs are considered a vital aspect to be diagnosed as soon as possible to minimize the risk. Hence, early prediction of CVDs needs to be examined in a patient (2, 3).

Cardiovascular disease comprises heart or cardiac diseases in heart patients and is a critical challenge in the medical field. Several reasons and findings that cause cardiac arrest are observed, such as a change in personal and professional lifestyles, habits, inactive lifestyle, growing age, habitual history related to smoking, alcohol consumption, stress level, and physiological signs, such as diabetes, high level of blood pressure (BP), obesity, cholesterol, hypertension, and existing heart problems. These risk factors need early, accurate, and efficient diagnosis to prevent cardiac arrests (4–6). The risk factors can be depreciated by executing proper lifestyle activities, such as lowering salt usage, absorbing a healthy diet, prohibiting alcohol and tobacco use, and regular physical exercise (7). Medical sectors use the abovementioned vital signs to produce consequential information from data. The vital signs are generally collected from publicly available datasets, wearable gadgets, medical hospitals, or sensors. Monitoring health parameters using Internet-of-Things (IoT) is a forming trend for future well-being. IoT sensors are mostly used to collect real-time vital signs and monitor the health parameters of individuals (8). Collecting, processing, and analyzing vitals help predict risk early to tackle the problem (9, 10).

One of the broad areas of Artificial Intelligence (AI) is machine learning (ML). ML is an effective technology and efficient field, which is based on prediction purposes. Using ML concepts, we can develop models or enable human abilities and train our data collected or gathered from the past for future predictive analysis (11). Those data are further divided into training and testing, which help in predicting future possibilities. This combined technology is called ML (12). The ML algorithm needs pertinent information for training and testing. In addition, the model performance can be raised if the balanced data are given to the ML model. In addition to that, the capability and accuracy of the predictive model can be improved if relevant features are selected from a dataset. Hence, a balanced dataset and feature selection are paramount for improving model performance (13). Furthermore, for cardiac arrest prediction, the ML algorithm is used, which has been addressed in this research. Nowadays, deep learning models show prominent growth and improvement in predicting and analyzing heart diseases (14).

Several studies have shown predictions on publicly available datasets. Still, this research aims to find early risk prediction of cardiac arrest based on real-time data collected from an individual using sensors and equipment. Furthermore, the authors build a Neural Network, Bagging, and Boosting model using ML techniques by selecting relevant datasets to show which model accomplishes best. This study also aims to classify risk in different age and gender groups of people. The points mentioned below show the main contributions of the article:

- To collect real-time vital signs of an individual using sensors and equipment.
- To detect cardiac arrest by applying the ML algorithm.
- To show a comparative analysis of the various ML classifiers.
- To classify cardiac arrest risk for the coming 10 years in different age groups.
- To show survival probability gender-wise in our creation of data.

The rest of the article is assembled as follows. Section Related Work narrates research work on heart diseases along with existing methods, datasets, and techniques available. Section Proposed Work exhibits the study of the proposed study of the author. Section Methodology reveals the methodologies and algorithms used for disease risk prediction. Section Risk Classification conveys the risk classification in people. Section Exploratory Data Analysis presents the results of Exploratory Data Analysis. Section Implementation Details communicates the implementation details of all the experiments and results. Finally, section Discussion, Conclusion, and Future Work discusses and concludes the research study of the author.

RELATED STUDY

Literature Survey

This section discusses the existing article to predict and diagnose heart diseases using several techniques and datasets used with various features and classification techniques.

Ambekar and Phalnikar (1) proposed a risk model for heart disease with structured data using a Convolutional Neural Network-based Unimodal Disease Risk Prediction (CNN-UDRP) model to classify risk levels as high, low, and medium. The dataset was extracted from the UCI repository. To avoid missing values, performed data cleaning and imputation to substitute missing values and increase the performance of the model were implemented. The proposed model achieves an accuracy of nearly about 65%. Ramprakash et al. (2) developed a model using a Deep Neural Network (DNN) and χ^2–a statistical method to predict the risk of heart patients. The Cleveland dataset was used, which is publicly available on the UCI repository. In this study, DNN models Artificial Neural Network (ANN) and DNN and proposed statistical methods χ^2-ANN and χ^2-DNN were compared. The comparison shows that the proposed method χ^2-DNN was more efficient in providing an accurate accuracy of 94%. Maini et al. (5) discuss data mining techniques to detect heart disease risk at an early stage and show the importance of assigning a cloud-based approach for helping healthcare providers. Mohan et al. (6) proposed a hybrid model combination of Linear model and Random Forest (RF) technique with different features combined on publicly available datasets, i.e., Cleveland from UCI repository and compared its result with models of Naïve Bayes (NB), RF, Linear model, Deep Learning, Logistic regression, etc. From all the proposed models, Random Forest with Linear Model (HRFLM) outperforms all others with 88.4% accuracy. Shah et al. (3) used the Cleveland dataset available from the UCI repository. They implemented ML algorithms, such as Decision Tree (DT), K-Nearest Neighbor (KNN), NB, and RF for visualizing the probability of heart disease in the patients. Among all models, KNN shows the highest accuracy score of 90% in predicting heart disease.

Garg et al. (15) used an ML algorithm to predict and classify diseases of a person who is suffering from heart problems. They have used a dataset from Kaggle, which was commonly used

for prediction purposes. From the attributes, 13 main attributes, such as the age of a person, chest pain, cholesterol level, and many more, were considered to predict heart diseases. The ML algorithms used are KNN and RF. The comparison was made between the ML algorithm and RF achieves the highest accuracy with 82%. Yadav et al. (16) proposed an optimization model, namely, "Optimized DNN using Talos" deploy ML classifiers, such as DT, KNN, RF, and Ensemble model (used ANN, KNN, and Support Vector Machine, SVM), for the prediction of heart diseases. They have used the concept of dimensional reduction where only vital information of patients was considered. On comparing their model with traditional models, the proposed model provides better accuracy and prediction.

Latha and Jeeva (17) convey classification as a commonly used technique in ML algorithms, so they have tried a new technique that is an ensemble technique that combines two or more classifiers to improve the accuracy obtained by a simple classification technique. They focus on increasing the accuracy and implementation of feature selection with a different set of features. There was a significant increment in the prediction accuracy. Singh et al. (18) proposed a multisurface proximal SVM (MPSVM)-based DT, collaborated with the ensembling method consisting of Gradient Boosting and RF algorithms on Cleveland dataset from UCI repository. They have solved the 2-class problem and the 5-class problem on a dataset using the proposed method. On analyzing, the 2-class problem shows better accuracy of 91% than the 5-class problem with 73%. Sowmiya and Sumitra (19) used the Cleveland dataset from the UCI repository to implement heart disease prediction. They have applied an ant colony optimization technique for the best feature selection for the hybrid KNN classifier. This hybrid model is compared to other classification algorithms, such as SVM, KNN, NB, C4.5, and DT. Their study proposed HKNN as an effective and efficient technique for heart disease prediction.

Alotaibi (20) focused on improving the performance and accuracy of ML algorithms compared with the previous studies. In this study, the prediction of heart diseases is made using RapidMiner tool on Cleveland dataset extracted from UCI repository and compared the results with previous work that used Weka and MATLAB tools where the performance of RapidMiner tool gives higher accuracy using techniques: SVM, DT, and Logistic regression.

Nikookar and Naderi (21) proposed a hybrid ensemble method to show comparison with the primary ensemble method using SPECT heart disease dataset, consists of SPECT images based on feature vector on fuser classifier algorithm, i.e., Adaboost, MLP, Logitboost, and RF. The comparison was made between hybrid and basic ensemble classifiers where the proposed hybrid classifier shows higher accuracy. Tama et al. (22) proposed a model using a two-tier ensemble technique to detect heart disease. The developed model was implemented by ensembling gradient boosting, RF, and extreme gradient boosting (XGBoost) classifiers. The model was trained and validated on available datasets: Cleveland, Statlog, Hungarian, and Z-Alizadeh Sani. A two-step significance test was conducted to compare with the existing article and provided the highest results. Li et al. (13) developed a diagnosis system called Fast Conditional Mutual Information (FCMIM-SVM) based on classification algorithms—ANN, SVM, and Logistic Regression—to diagnose heart diseases using the Cleveland dataset. They used minimal redundancy maximal relevance, relief, most minor absolute shrinkage selector operator, etc., as a feature selection algorithm to select relevant features to get accurate accuracy. The proposed model shows a better accuracy as compared to the previously developed model. Sarmah (8) proposed a model using IOT sensors attached to the body of the patient to gather real-time data and applied the Deep learning modified neural network (DLMNN) model for heart disease prediction. The prediction from the model executes in three ways: Authentication, Encryption, and Classification. Thus, the classification shows normal and abnormal output and prescribes it in its manner. Hence, the model proposed shows improvement over the existing algorithm with providing authentication.

Chauhan et al. (10) used an ML algorithm to predict heart diseases on the Cleveland dataset publicly available and compared the result with the algorithms. Among all, ANN outperforms all classification algorithms with 85% accuracy. Pan et al. (14) proposed an enhanced deep-learning Convolution Neural Network (CNN) model that predicts and improves heart disease in a patient. This model uses the more profound architecture of the multilayer perceptron model and regularizes parameters. In addition, this system has been executed on the Internet-of-Medical-Things (IOMT) platform to provide efficient solutions for a doctor for the diagnosis of heart diseases. This CNN model is compared with other ML techniques and shows a maximum accuracy of 97%. Fitriyani et al. (7) developed a clinical decision support system (CDSS) to recognize early heart failure. They also proposed the Heart Disease Prediction Model (HDPM) for the CDSS system to eliminate noise and outlier presence. They have built the Starlog and Cleveland dataset model, which is available publicly and tested on other ML models and compared the results. The proposed model achieves 95% with Statlog and 98% with the Cleveland dataset.

Tate and Rao (23) give an idea of using wearables in form of a wristwatch, wrist band, ambulatory devices, skin response monitors, Fitbit tracker with an app installed, and many more for prediction of cardiac arrest at an early stage. These electronic gadgets help in tracking our day-to-day routine and monitoring our daily activities by keeping track of each activity. The vital information of patients can assist individuals by sending emergency alerts via web and mobile interfaces. Several studies, such as Framingham Heart Study, had developed the Cox regression hazard ratio model specifically for the sex groups for the prediction of heart diseases by estimating 10-year risk. Hence, the idea of capturing the vitals of an individual from wearables helps to get a better prediction and classification rates for detecting heart diseases. Prabhu et al. (24) proposed the CN-Based Multimodal Disease Risk Prediction (CNN-MDRP) model for risk prediction of diseases on a sizeable Medical dataset and compared the result with the algorithm that is CNN-UDRP. The performance of CNN-MDRP shows more improvement than the CNN-UDRP model for risk prediction. Shankar et al. (25) aim to predict heart diseases in an individual. They have made the

prediction based on details entered by a patient in a hospital, and the data were trained on a CNN model to find the accuracy. The data used were in structured and unstructured formats. The proposed CNN model shows improvement by comparing it with other algorithms.

Singh and Kumar (12) have calculated accuracy for predicting heart diseases using ML algorithms. The algorithms were KNN, DT, Linear Regression, and SVM utilizing Cleveland dataset from the UCI repository for training and testing. Overall, the accuracy of the KNN algorithm achieves 87% for early heart disease prediction. Amin et al. (26) have collected real-time data, i.e., the vital signs of a patient using Fitbit wearable. The data captured by the wearable were entered into an application and processed through an ML algorithm. This study aims to notify an individual whether he/she is at risk of CVD or not. The outcome of the prediction result helps doctors in diagnosis patient health. Chen et al. (27) proposed a technique for predicting the occurrence of perioperative heart diseases. The data were collected from the hospital and divided into structured data in numerical form and unstructured data in textual form. ML algorithms for numerical data gradient boosting tree algorithm and textual data topic model of text-based data model were used. The datasets were fused using a simple logistic regression model and achieved a sensitivity of 90% and specificity of 93%.

Cohen et al. (28) proposed a 1-year risk prediction model for patients with congenital heart diseases. The dataset was collected from Quebec hospitals of patients varying in age between 18 and 64 years. Multivariate logistic regression was used to predict the risk for heart diseases. The risk score was indicated between 0 and 19. Hence, the risk score model shows excellent performance. Chang et al. (29) use an ML classification algorithm to predict the vital signs of a patient for the next hour. Vital signs, such as heart rate (HR), BP, mean arterial pressure, and oxygen, were considered. Early warning signs can help nurses and doctors treat the patient before the situation degrades. ML algorithms, such as ANN, RF, Gradient Boosting, and long-short-term-memory (LSTM), were developed. The outcome shows the accuracy of ML models. Youssef Ali Amer et al. (30) have carried our experiment with vital signs of patients using wearable devices. The vital signs considered were BP, oxygen level, and HR. These vital signs were estimated every minute using mean and minimum statistical values.

Then, the ML technique of KNN LS SVM was executed to predict future values of vital signs that were monitored. The performance of implemented ML algorithm was compared with the LSTM approach, and the proposed approach shows much improvement in predicting the early warning score of vital signs. Stehlik et al. (31) have indicated re-hospitalization of HF by non-invasive and remote monitoring. Wearable sensors were placed on the chest of a person (up to 3 months), collecting and recording physiological data (vital signs). The data were uploaded on a cloud platform via a smartphone. Analytics were performed on data to detect HF exacerbation. The result shows 76–88% sensitivity and 85% specificity. An alert was shown 6 days prior for re-admission.

Jia et al. (32) use the Cox proportional hazards regression model to find the probability of risk for 10 years using the Framingham dataset. The risk factors considered were sex, body mass index (BMI), systolic BP (SBP), and diabetes. They have calculated a sex-based risk score using the Cox regression hazard equation.

Yang et al. (33) proposed a multivariate regression model for the prediction of CVDs. The prediction was made on the collection of vital information of patients from the centers of the national high-risk program by having a cardiac events assessment. The dataset consists of all information, such as BP, cholesterol level, obesity, smoking, BMI, and many more attributes were taken into consideration. The area under curve results of the multivariate regression model were then compared to other ML classification techniques, such as CART, NB, and Ada Boost. The outcome result of the multivariate regression model outperforms other methods. Made et al. (34) surveyed CVDs estimating 10-year risk in the different age groups of people. They proposed a SCORE model that uses risk factors, such as weight, BP, cholesterol level, smoking habit, diabetes, and many more. The data were collected by taking samples of men and women of age 18 and above. The 10-year risk estimation was made on statistical analysis, which estimates coronary heart diseases and no-coronary heart disease. These studies on 10-year risk scores help to prevent heart diseases in the coming future.

Comparison of Previous Findings and Proposed Study

The literature study discussed above and in **Table 1** shows the importance of the vital signs of a person in the healthcare industry to detect health-related problems. Mainly, health problems are majorly concerned with the heart, such as cardiac arrest, CVDs, coronary heart disease, and many more. In this research, the authors have depicted the detection of cardiac arrest in a person by collecting real-time data. Moreover, most of the detection and prediction of heart disease is majorly made on a public dataset consisting of shared attributes. In our contribution, the features chosen are mainly focused on cardiac arrest detection. After detecting cardiac arrest, gender-wise survival probability, and age-wise cardiac arrest risk scores are also formulated using two approaches: (1) Kaplan-Meier and (2) Cox regression model. These two methods help to provide which gender has fewer survival chances and which age group has a high probability of risk in the next 10 years.

PROPOSED WORK

This section briefly describes the novelty of the proposed work (contributions) and the articles mentioned early by some research on the same Cox regression model.

The progress in CVD risk models is ordinarily based on conventional laboratory-based predictors. The primary purpose was to develop a prediction model, which was based on a 10-year risk. To predict 10-year risk, a Cox regression proportional method was applied. A dataset from Framingham Original Cohort and SCORE of men and women aged between 30 and 62 was taken to predicate CVD risk. In this study, the CVD risk was

TABLE 1 | Comparison of literature work and proposed work.

References	Findings	Method used	Results obtained
Ramprakash et al. (2)	Prediction of heart disease on Cleveland dataset	Machine learning algorithms—Deep Neural Network and χ^2 —a statistical method	Accuracy—94%
Mohan et al. (6)	Prediction of heart disease on Cleveland dataset	Proposed HRFLM model using Random Forest, and Linear model	Proposed model accuracy—88%
Sowmiya and Sumitra (19)	Used Cleveland dataset to implement heart disease prediction	Used ant colony optimization technique for hybrid KNN classifier and compared with other ML models	Proposed approach HKNN proves efficient and effective technique
Made et al. (34)	Estimation of 10-year coronary heart risk	Using statistical approach—regression analysis in age group of 18 and above	Study estimates heart risk for the next 10-years
Proposed work	Detection and classification of cardiovascular diseases in different age-based group and gender-based people, collected vital signs of an individual using wearables and medical equipments	Used 6 Machine Learning algorithm—Decision Tree, SVM, ANN, RFC, XGBoost, and NB. Further, comparison is made among 6 classifiers. Also, we developed two models based on Kaplan-Meier method (For Gender-Based) and Cox-Regression Proportional Model (For Age-Based) using scoresheet	Among 6 Machine learning models-Decision Tree achieves maximum accuracy of 98%. Our proposed approach, for age-based and gender-based shows risk score (in percentage) of cardiac arrest for the next 10-years

being predicted on a gender basis (men/woman). The risk factors involved are age, sex, BMI, SBP, and diabetes.

A proposed study related to cardiac arrest prediction using the real-time dataset is classified based on gender and age. Our study has used two methods to find the risk classification probabilities among gender-based and age-based. The two methods used—Kaplan-Meier survival analysis and Cox regression model survival analysis—are considered. These two methods have their importance in analyzing risk probabilities.

Kaplan-Meier Method

This method is a non-parametric survival function for the estimation of survival probability. Kaplan-Meier method also estimates the survival curve by giving a statistical comparison between two groups: men and women.

Cox Regression Proportional Model

Cox regression proportional model does regression analysis with survival data without making any strong assumption. It calculates the hazard ratio (HR) of the covariates used in our data and is based on the equation, and it shows risk probability. Cox regression proportional model is a model that helps to find risk probability or score on more than one covariate. In general, it is a method to identify the effect of variables at some time interval on some event that occurs.

METHODOLOGY

Measuring Device

The risk factors associated with heart diseases are cardiac arrest prediction: age, gender, height, weight, BMI, BP [SBP and diastolic BP (DBP)], HR (beats per minute), and oxygen level (SPO_2%).

The above factors used in this research are collected using IOT sensors and equipment for different risk factors. The measuring device can capture the vital signs of a person needed to estimate cardiac arrest prediction. **Figures 1**–**3** show measuring device used for capturing vital signs.

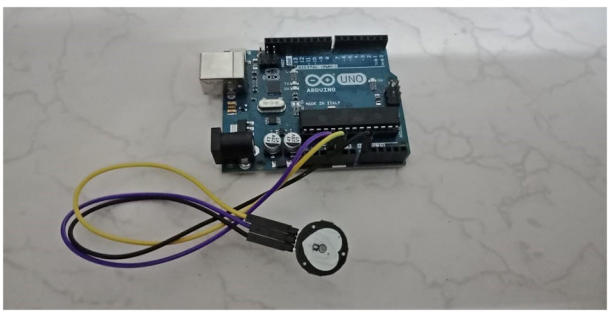


FIGURE 1 | Pulse sensor connected with Arduino.

FIGURE 2 | Sphygmomanometer.

FIGURE 3 | Oximeter.

Data Collection

Pulse Sensor

A pulse sensor is a device that is used to detect biometric pulse rate or HR, and it also keeps track of volume change. It is generally helpful for healthcare applications to capture the pulse rate of an individual. A pulse rate is detected when there is a change in blood vessel volume, and that occurs when the heart pumps blood.

For this research, we used a pulse sensor that is connected to Arduino UNO Board. Above, **Figure 1** shows the configuration of the pulse sensor and Arduino Board. A pulse rate or HR was captured from a total of 18 patients in beats per minute (BPM) at a 2 min interval.

Sphygmomanometer

A sphygmomanometer is a device used to measure the BP of a person. The connected rubber cuff with the device is wrapped around the upper arm of a patient and records the blood pressure in SBP and DBP.

For this research, we used Sphygmomanometer to measure BP. Above, **Figure 2** shows a Sphygmomanometer connected to the arm of a patient. BP was captured from the same 18 patients at every 2 min interval.

Pulse Oximeter

The pulse oximeter is a measuring device used to keep track of oxygen levels in a human body. This is an easy and painless measuring device to see how much oxygen reaches the different parts of the body. This device is generally used in hospitals and at home to determine the oxygen level in the body. It measures the oxygen level in SPO_2% form.

For this research, we used a pulse oximeter to measure the SPO_2. Above, **Figure 3** shows a pulse oximeter placed on the fingertip of a patient and the estimated oxygen level. It is a non-invasive method for monitoring oxygen saturation. Oxygen level was captured from the same 18 patients at every 2 min interval.

Dataset Description

Table 2 shows the names of attributes along with their feature description. The table also contains a range column indicating the range value mentioned in our dataset and its dataset. The dataset comprised attributes are:

- Date and time
- Gender
- Age
- Height
- Weight
- BMI
- SBP
- DBP
- HR (BPM)
- Cardiac arrest prediction—target variable.

Cardiac arrest prediction is our target variable from the features mentioned above, which is predicted in our dataset.

The target variable—cardiac arrest prediction—is labeled using the following conditions, as mentioned in **Table 3**.

Data Pre-processing

In the ML process, the processing of datasets is essential for good impersonation. Some techniques for data pre-processing are data cleaning, data integration, data transformation, data reduction, and data discretization. In our dataset, there are 540 records where the column gender is denoted as "F" and "M." These need to be processed as "0" and "1" using LabelEncoder.

Algorithm Used

Artificial Neural Network

Artificial Neural Network consists of ample neurons simulated together to form a network architecture. This architecture of the neural network is used to transform the inputs into some meaningful outputs. ANN is a supervised ML technique generally used for prediction. ANN consists of an input layer, hidden layer, and output layer linked together, forming a network. These layers are associated with numerical weights to minimize the error mathematically. For the cardiac arrest prediction dataset, the input layer consists of SBP, DBP, HR, and oxygen levels, fed to 16 neurons. These input parameters are fed into the next layer. The hidden layer consists of eight neurons and finally, after the processes, gives output to the output layer as a final result.

Random Forest Classifier (RFC)

Random Forest Classifier is a type of supervised ML algorithm used for two purposes: classification and regression problems.

TABLE 2 | Dataset description.

Sr. No.	Attributes name	Description	Value range	Type
1	Date and time	Shows person data collected on date and time	–	DD-MM-YYYY hh:mm:ss
2	Gender	Shows gender of a person	Male = 1 Female = 0	Categorical
3	Age	Shows age of a person in years	Age from 15 to 75	Numerical
4	Height	Shows height of a person in feet and inches	Height varies from 4′3′′-5′9′′	Numerical (in feet and inches)
5	Weight	Shows weight of a person in kilogram (kg)	Weight ranges from 45 to 75 kg	Numerical (in kg)
6	Body Mass Index (BMI)	Shows body mass of a person	BMI ranges from 15.1 to 38.7	Float
7	Systolic blood pressure (BP)	Shows systolic blood pressure of a person in mmHg	Normal Range must be ≥120 to ≤134 mmHg	Numeric
8	Diastolic blood pressure (BP)	Shows diastolic blood pressure of a person in mmHg	Normal range must be ≥79 to ≤87	Numeric
9	Heart rate (BPM)	Shows heart rate of a person in beats per minute	Heart rate between ≥49 to ≤84	Numeric
10	Oxygen level (SPO_2)	Shows oxygen level in a person body	Range between ≥95% to ≤100%	Numeric in %
11	Cardiac arrest prediction	Target variable to be predicted	0 = No cardiac arrest 1 = Cardiac arrest	Binary value

TABLE 3 | Labeled target variable.

Parameter	"0"—predicted label	"1"—predicted label
Systolic blood pressure mmHg	Between ≥120 to ≤134 mmHg	Below 120 mmHg and above 134 mmHg
Diastolic blood pressure mmHg	Between ≥79 to ≤87 mmHg	Below 79 mmHg and above 87 mmHg
Heart rate (BPM)	Between ≥49 to ≤84	Below 49 and above 84
Oxygen level (SPO_2)	Between 95–100%	Below 95% or above 100%

It follows the concept of ensemble learning, which combines multiple classifiers to solve complex problems and improves the performance of a model. Generally, it consists of several DTs forming a subset of a given dataset and improving the predictive accuracy of a dataset. RFC predicts the final output by comparing the prediction from each tree and based on the decision, such as majority voting; it indicates the result. For the cardiac arrest prediction dataset, the input parameters considered are SBP, DBP, HR, and oxygen level. The dataset is divided into 80% training and 20% testing set and selected the number of trees estimator 30 and function "entropy" for the information gain.

Gradient Boosting (XGBoost)

"XGBoost" stands for extreme gradient boosting and can be implemented with the gradient boosted trees algorithm. XGBoost is a supervised ML algorithm and is considered one of the most sought-after ML techniques. This is due to its high prediction accuracy and user-friendliness. Just like other ML algorithms, XGBoost can also be used for regression and classification of problems. XGBoost uses DTs as a weak predictor and minimizes the loss function using a gradient descent algorithm. In implementation, we created an "XGB Classifier" object by passing parameters subsample = 0.7, max_depth = 5, and n_estimators = 100 and by default it takes booster = "gbtree." The dataset is divided into 80% training and 20% testing set where training data are used to train a model and test part to measure its performance.

Support Vector Machine

Support Vector Machine is the best-supervised ML algorithm. It is used for both regressions and classification problems. However, for the dataset used, this algorithm is used for classification problems. For the cardiac arrest prediction dataset, the input parameters considered are SBP, DBP, HR, and oxygen level. The data are divided into 80% training and 20% testing set. We have used the Kernal value "rbf" Radial Basis Function for a non-linear classification problem.

Naïve Bayes

Naïve Bayes is a supervised ML algorithm that is used for solving a classification problem. It is generally based on the Bayes theorem and is considered one of the best classification techniques. It makes a quick prediction based on object probability. It builds a model on the classification problem and shows the high accurate result for a large dataset. For the cardiac arrest prediction dataset, the input parameters considered are SBP, DBP, HR, and oxygen level. The data are split into 80% training and 20% testing set. The model used in making a prediction is the Gaussian NB model, which follows a normal distribution for our predictor values.

Decision Tree

A DT is a classification technique used to solve a classification problem. It is a simple and widely used classification method. It is generated by providing a set of inputs in the form of a dataset. Its structure is like a flowchart that consists of Root, Sub-tree, and Leaf nodes. The algorithm starts from the root node, compares the values with the dataset attribute, and on the comparison, it follows the tree structure, i.e., branch and jumps to the next node. It continues its process until it reaches the last node, i.e., a leaf node. This helps in predicting the class of the given dataset. For the cardiac arrest prediction dataset, the input

parameters considered are SBP, DBP, HR, and oxygen level. The data are split into 80% training and 20% testing set, considering the maximum depth of tree-level to be three and attribute used "Gini" to measure purity while creating a DT.

RISK CLASSIFICATION

Survival analysis is a statistical approach for analyzing data where the outcome variable is the time till an event occurs. They are categorized into parametric, semi-parametric, and non-parametric models. For analyzing data, several survival analysis techniques are used, such as Life Tables, Kaplan-Meier analysis, Cox proportional hazards regression model, Survival trees, and Survival random forest. In this section, the different methods we used for our survival analysis are introduced.

Risk Survival Analysis Model

In this study, two models are proposed: Age_Cox_model specific for age-based survival analysis and Gender_KM_model for gender-based survival analysis. Both models are used to find probability risk in an individual. Age_Cox_model is generally based on the Cox regression model to predict the survival probability of having different ages. Gender_KM_model is based on the Kaplan-Meier estimator to analyze the survival probability based on male and female groups.

Age-Wise Classification

The age-based classification is done using Age_Cox_model analysis. The first stage of Age_Cox_model is to provide input in the form of a dataset. The next step is the pre-process the dataset to solve the categorical data such as gender. For categorical problems, the use of LabelEncoder is applied for the gender feature. The third stage is to select relevant parts for our proposed model. After selecting features, the next step is to train our model by applying the Cox regression model. The model then predicts the result, showing the importance of features and whether there are any risks. The steps of our model are described in the following pseudocode.

Step 1: Import libraries,
Step 2: Load the dataset,
Step 3: Pre-processed the data,
Step 4: Select relevant features,
Step 5: Apply Cox regression CoxFit() model,
Step 6: Predict results show feature importance and chances of risk.

The flow of the work is presented in the below diagram, **Figure 4**.

The above diagram result shows the importance and risk probability of features. **Table 4** describes that the *p*-value must be <0.05 and is considered significant.

The *p*-value of attributes age, SBP, and HR (BPM) is considered a significant covariant as values are <0.05.

On considering HR, i.e., exp(coef) shows the level of risk and have the following condition:

If HR = 1 Covariant has no effect of risk
HR <1 Covariant has fewer chances of risk
HR >1 Covariant has more options of risk.

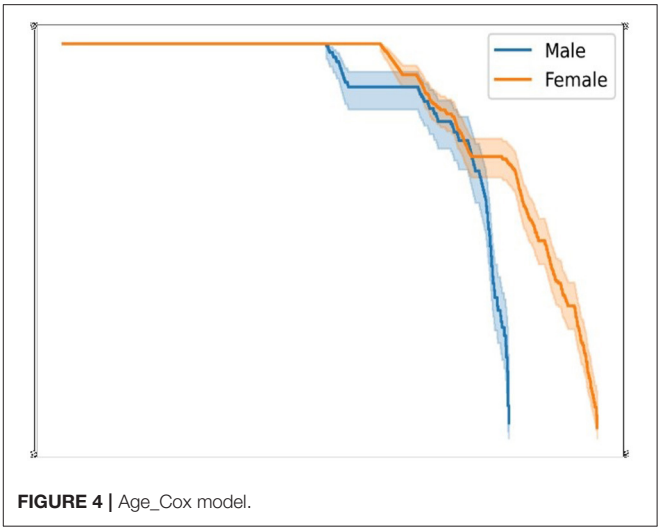


FIGURE 4 | Age_Cox model.

Hence, to know the chances of risk in our proposed work, age-based classification can be solved using the Cox regression survival analysis model formula, which is explained in the next part.

A. Age-based risk score calculation

The age-wise classification is derived from the prediction of a 10-year risk score for cardiac arrest prediction. Among the features collected, the main risk factors for cardiac arrest prediction are a Person's age, BP (SBP and DBP), HR (measured in BPM), cardiac arrest prediction as target variable, and T1 factor as a time interval. The age-wise classification for 10-year risk probability is estimated using the Cox regression model.

Table 5 represents the characteristics of risk factors that show statistical values of "total number of observations" (in each column), "number of observations with missing values," "number of observations without missing values," "minimum value" (in each variable column), "maximum value" (in each variable column), "mean value" (in each variable column), and "SD value" (in each variable column) are encapsulated.

Table 6 shows the summary statistics of events that occur while modeling risk prediction for 10 years. It shows, "total observed" = 540 total number of data or observation, "total failed" = 328 denotes the number of possible cardiac arrests, "total censored" = 212 denotes thin chances of cardiac arrest, and "Timestep" = 540 denotes the number of time steps equals to the total number of observations.

The following table indicator, **Table 7**, represents the model quality, that is, "efficiency of the model/variable fit." The most important statistic value to be noted is the probability of the Chi-square test on the log-ratio as shown in **Table 8**; on comparing, the model statistics with the defined covariates, the probability value, that is, the *p*-value must be lower than 0.05 ($p < 0.05$), to be significant. Hence, from **Table 7**, we conclude that the variables show significant information.

The following table gives information on the model. **Table 9** helps us understand the effect of our risk factors on the model. From **Table 9**, as per our collected dataset, the variable that

TABLE 4 | Summary of Age_Cox_model.

	β	Exp(β)	Se(β)	β lower 95%	β upper 95%	Exp(β) lower 95%	Exp(β) upper 95%	z	p	$-\log_2(p)$
Age	0.03	1.03	0.00	0.02	0.03	1.02	1.03	7.32	<0.005	41.86
Systolic BP	−0.04	0.96	0.01	−0.05	−0.03	0.95	0.97	−7.35	<0.005	42.25
Diastolic BP	0.01	1.01	0.01	−0.01	0.02	0.99	1.02	0.72	0.47	1.09
Heart rate (BPM)	0.00	1.00	0.00	0.00	0.01	1.00	1.01	2.88	<0.005	7.98

TABLE 5 | Summary of statistical values of risk factors used for 10-year risk model.

Feature	Obs.	Min	Max	Mean	Std. deviation
T1	540	672.00	1363.00	1076.831	167.757
Cardiac arrest prediction	540	0.00	1.00	0.607	0.489
Age	540	15.00	75.00	43.056	19.364
Systolic BP	540	82.00	157.00	118.961	14.896
Diastolic BP	540	55.00	125.00	78.113	9.905
Heart rate (BPM)	540	24.00	238.00	87.006	39.929

TABLE 6 | Summary statistics (events).

Total observed	Total failed	Total censored	Time steps
540	328	212	540

TABLE 7 | The goodness of fit statistic.

Statistic	Independent	Full
Observations	328.000	328.000
DF	0.000	4.000
−2 Log (likelihood)	3487.800	3398.055
AIC	3487.800	3406.055
SBC	3487.800	3421.227
Iterations	1.000	3.000

TABLE 8 | Test of the null hypothesis H0: beta = 0.

Statistic	DF	Chi-square	Pr > Chi^2
−2 Log (likelihood)	4	89.74558	<0.0001
Score	4	94.97911	<0.0001
Wald	4	94.87288	<0.0001

influences our risk survival model is SBP. From the beginning of the study, this risk variable indicates the patient has a high effect on cardiac arrest prediction. In this study, from the table, the "Value" column shows the HR, which is obtained as the estimated parameter exponent and is calculated using Equation (1).

The HR is given using the mentioned formula:

$$H(t) = h_0(t)\exp(\sum ßi X_i)(i = 1 \text{ to } r) \quad (1) \tag{1}$$

Notations are:

t = Survival time (10-year risk) (t = 10).
exp = Exponential function (exp(X)).
B_i = It measures the impact of covariates (B_1, B_2, B_3, …, B_r).
X_i = Predictor variables (X_1, X_2, X_3, …, X_r).
$h_0(t)$ = Baseline hazard rate.
H(t) = HR.

The Cox model regression has an exponential form, as seen in Equation (1).

As per our dataset, for 10 years, the term $h_0(t)$ = baseline survival rate value is considered 0.993 at mean values of covariates from the survival distribution function. Here, "t" represents an event that occurred at a time. The H(t) hazard function is determined by the set of covariates X_i = (X_1, X_2, X_3, …, X_r) and the regression coefficient $B_{i=}$ (B_1, B_2, B_3, …, B_r) that measures the impact of covariates.

In **Table 10**, the column "rho," correlates between residuals and time vector as in our above case—Kaplan Meier. The other two columns, "Chi-square" and "Pr > Chi-square," show test statistics and the associated *p*-value. Here, the *p*-values must be <0.005, and the value from **Table 10** reveals no violation of the risk assumption.

Hence, to find the survival risk score from the Cox model regression, it can be written as:

$$R(t) = 1 - R0(t)\exp(\sum ßiXi - \sum ßi\overline{X}i)\ (i = 1 \text{ to } r) \quad (2) \tag{2}$$

R(t) is the risk estimation of cardiac arrest of an individual; $R_0(t)$ is the baseline survival time at t; βi is the regression coefficient covariates values mentioned in **Table 8**; X_i is the value of the i_{th}: the risk factor value [log-transformed value – natural log (ln)]; and $\overline{X}_I$is the mean value of the covariates. Here, the i value ranges from 1 to r and denotes the risk factors considered. The cardiac arrest risk score or percent can be determined using the mentioned formula as shown in Equation (2).

The estimation of the Cox regression model is formulated from the scoresheet, and the Cox regression equation is derived. **Tables 11**, **12** show the scoresheet that represents the points allocated

TABLE 9 | Regression coefficients.

Variable	Value	Std. error	Wald Chi-square	Pr > Chi^2	HR	HR < 95%	HR > 95%
Age	0.027	0.004	53.587	<0.0001	1.027	1.020	1.035
Systolic BP	−0.039	0.005	54.068	<0.0001	0.962	0.000	0.972
Diastolic BP	0.005	0.007	0.517	0.472	1.005	0.991	1.019
Heart rate (BPM)	0.004	0.001	8.232	0.004	1.004	1.001	1.007

TABLE 10 | Proportionality test.

Variable	Rho	Chi-square	Pr > Chi^2
Age	−0.22834	24.24338	<0.0001
Systolic BP	0.09274	4.133863	0.042
Diastolic BP	0.012658	0.061134	0.805
Heart rate (BPM)	0.136702	7.842881	0.005
Global		32.0064	<0.0001

TABLE 11 | Cardiac arrest risk score for men.

Male			
Points	**Age**	**Blood pressure**	**Heart rate**
−1	30–34		<50
0	35–39	<130/85	50–70
1	40–44	130/85–139/89	70–90
2	45–49	140/90–159/99	90+
3	50–54	≥160/100	
4	55–59		
5	60–64		
6	65–69		
7	70–74		

TABLE 12 | Cardiac arrest risk score for women.

Female			
Points	**Age**	**Blood pressure**	**Heart rate**
−9	30–34		<50
−4	35–39	<120/89	50–70
0	40–44	120/80–139/89	70–90
3	45–49	140/90–159/99	90+
6	50–54	≥160/100	
7	55–59		
8	60–74		

to an individual risk factor for both genders. **Table 13** shows the total risk score probability after formulating Equation (2).

The calculation for estimation of cardiac arrest risk is derived below:

TABLE 13 | Ten-year risk for the Risk points calculated.

10-year risk score (%)			
Points	**Male**	**Points**	**Female**
−3	1%	<-2	1%
−2 or −1	2%	−1, 0, or 1	2%
0	3%	2 or 3	3%
1 or 2	4%	4	4%
3	6%	5	5%
4	7%	6%	6%
5	9%	7	7%
6	11%	8	8%
7	14%	9	9%
8	18%	10	11%
9	22%	11	13%
10	27%	12	15%
11	33%	13	17%
12	40%	14	20%
13	47%	15	24%
14 or more	56%	16	27%
		17 or more	32%

TABLE 14 | Risk point calculation.

Risk factor involved	Values	Points allocated
Gender	Female	–
Age	50	6
Systolic blood pressure	131	0
Diastolic blood pressure	80	0
Heart rate	112	3
Total points scored		9
Risk estimation according to sheet		9%

B. Risk estimation from the scoresheet

As mentioned in **Tables 11**, **12**, the risk points are calculated according to the covariate's ranges. **Table 14** shows points allocated according to the risk factor involved.

C. Risk estimation from Cox regression model

As per Equation (2), risk estimation of cardiac arrest in female is given by:

Formula:

$$R(t) = 1 - R0(t) \exp\left(\sum ßiXi - \sum ßi\overline{Xi}\right) (i = 1 \text{ to } r)$$

Values for:

- $R_0(t)$ for t = 10 years = 0.993 is the baseline survival
- βi is the estimated regression coefficient (in **Table 9**)
- Xi is the log-transformed value of ith risk factor
- $\overline{X}I$ is the mean value of the covariates considered (in **Table 5**)
- R is the number of risk factors involved in predicting risk. The risk estimation of the cardiac arrest based on the Cox regression model is computed as:
- Xi is the log-transformed value of ith risk factor
- $\overline{X}I$ is the mean value of the covariates considered (in **Table 5**)
- R is the number of risk factors involved in predicting risk.

The risk estimation of the cardiac arrest based on the Cox regression model is computed as:

$$\begin{aligned}\sum_{i=1}^{r}\beta iXi &= [0.027^{*}\ln(50)] + [-0.039^{*}\ln\ (131)] \\ &+ [0.005^{*}\ln(80)] + [0.004^{*}ln\ (112)] \\ &= 0.1056 - -0.1901 + 0.0219 + \ 0.0188 \\ &= -0.0438 \\ \sum_{i=1}^{r}\beta i\overline{X}i &= [0.027^{*}43.056] + [-0.039^{*}118.961] \\ &+ [0.005^{*}78.113] + [0.004^{*}\ 87.006] \\ &= 1.1625 - 4.6394 + 0.3905 + \ 0.3480 \\ &= -2.7384\end{aligned}$$

Substituting values in Equation (2) we get,

$$\begin{aligned}R(t) &= 1 - R_0(t)\exp(\sum_{i=1}^{r}\beta iXi - \sum_{i=\ 1}^{r}\beta i\overline{X}i) \\ &= 1 - \mathrm{R}_0(10)\exp(-0.0438 - \ (-2.7384)) \\ &= 1 - (0.993)\exp\ (2.6946) \\ &= 1 - (0.993)^{\ (14.79)} \\ &= 1 - \ 0.9013 \\ &= 0.0987 \\ &= 0.0987^{*}100(\text{convertingto \%}) \\ &= 9.87 \\ &\approx 9\%\end{aligned}$$

According to the scoresheet, the point system is 9%, and the Cox regression model also gives the approximate percentage of risk 9%. Therefore, a female having age 50 may have a chance of 9% more risk of cardiac arrest in the coming 10-year risk prediction.

Gender-Wise Classification

The gender-based classification is done using Gender_KM_model analysis. The first stage of Gender_KM_model is to provide input in the form of a dataset. The next step is to pre-process the dataset to solve the categorical data such as gender. For categorical problems, the use of LabelEncoder is applied for the gender feature. We divide our dataset into two groups: male and female and apply the Kaplan-Meier analysis model. The model then predicts the result, showing prediction probability in each gender at an event time. The steps of our model are described in the following pseudocode.

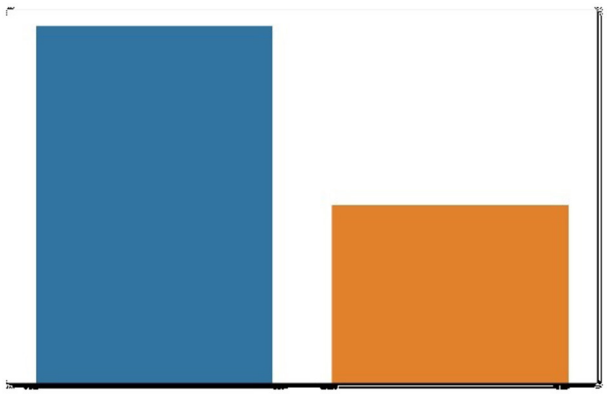


FIGURE 5 | Gender_KM model.

Step 1: Import libraries
Step 2: Load the dataset
Step 3: Pre-processed the data
Step 4: Divide dataset into two groups: male and female
Step 5: Apply Kaplan-Meier KaplanMeierFitter() model
Step 6: Predict results shows probability in each gender at an event time.

The flow of the work is presented in the below diagram, **Figure 5**.

A. Gender-based estimation of risk probability

The gender-wise risk classification can be estimated using a non-parametric survival estimator called the Kaplan-Meier method. It is useful in determining the survival probabilities of an individual and gender-wise. As per our collected dataset, survival function estimator S(t) can be given as:

$$\mathrm{S(t)} = \Pi(1 - \mathrm{di/ni})(\mathrm{i} : \mathrm{ti} <= \mathrm{t}) \tag{3}$$

Where,

S(t) = gives survival probability
t_i = time at when an event occurs
d_i = number of events occur (e.g., at risk of cardiac arrest) at time ti
n_i = number of individuals who have survived (e.g., not at risk of cardiac arrest) at time t_i

The Kaplan-Meier estimator helps to find the probability or survival rates of adjusted covariates. In our dataset, our estimation is to classify risk according to gender (male and female). We have divided our data into two groups: male and female. Our target is to notice any significant difference in probability or survival rate if data are divided into two groups based on gender. Based on gender split, we come up with survival probability in men and women based on Equation (3), which is shown in **Tables 15**, **16**.

The survival probability/risk of an individual (male/female) is estimated based on the time interval. On predicting survival probabilities in both genders using object creation of Kaplan-Meier we get,

TABLE 15 | Event table for male.

Event_at	Removed	Observed (d_i)	Censored	Entrance	At_risk (n_i)
0	0	0	0	180	180
672	1	1	0	0	180
674	1	0	1	0	179
676	1	1	0	0	178
678	1	1	0	0	177
680	1	1	0	0	176
682	1	0	1	0	175
684	1	0	1	0	174
686	1	0	1	0	173
688	1	1	0	0	172
690	1	1	0	0	171
692	1	0	1	0	170
694	1	0	1	0	169
696	1	1	0	0	168
698	1	1	0	0	167
.....	.....	.....	.....	.....	.....
1,120	2	1	1	0	20
1,122	2	0	2	0	18
1,124	2	0	2	0	16
1,126	2	1	1	0	14
1,128	2	0	2	0	12
1,130	2	1	1	0	10
1,132	2	1	1	0	8
1,134	2	1	1	0	6
1,136	2	2	0	0	4
1,138	2	1	1	0	2

TABLE 16 | Event table for female.

Event_at	Removed	Observed (d_i)	Censored	Entrance	At_risk (n_i)
0	0	0	0	360	360
810	1	1	0	0	360
812	1	1	0	0	359
814	1	1	0	0	358
816	1	1	0	0	357
818	1	1	0	0	356
820	1	1	0	0	355
822	1	1	0	0	354
824	1	0	1	0	353
826	1	1	0	0	352
828	1	1	0	0	351
830	1	1	0	0	350
832	1	1	0	0	349
.....	.....	.....	.....	.....	.....
1,345	2	1	1	0	20
1,347	2	2	0	0	18
1,349	2	1	1	0	16
1,351	2	1	1	0	14
1,353	2	1	1	0	12
1,355	2	1	1	0	10
1,357	2	1	1	0	8
1,359	2	0	2	0	6
1,361	2	1	1	0	4
1,363	2	1	1	0	2

kmf_m.predict (900) = 0.89061… ≈ 89%
kmf_f.predict (900) = 0.92205… ≈ 92%

Here, predict (900) indicates predicting survival or risk probability at 900-time intervals. The male probability of surviving is 89%, and the female probability of surviving is 92%.

From the above observation, we can conclude that the survival probability of women is higher compared to men. Men are at higher risk than women. Below, **Figure 6** shows the survival probability of men and women at a given time t.

EXPLORATORY DATA ANALYSIS

Figure 7 shows the number of female and male data in our dataset from the figure. We can conclude that there are 150 "men" and 350 "women."

Figure 8 shows a correlation graph among attributes, using the syntax:

data.corr( )

Figure 9 shows counts of cardiac arrest prediction "1" and "0" in each gender, male and female. We conclude that 100 "men" are prone to cardiac arrest, and around 75 "men" are risk free of cardiac arrest. Similarly, more than 200 "women" are prone to cardiac arrest, and ∼140 "women" are risk free from cardiac arrest.

Figure 10 shows the various age distributions. It is observed from the bar plot that 25 years of people are seen more following 17 and 63 years of people are found more.

Figure 11 shows the distribution of age, and it seems to be normally distributed.

IMPLEMENTATION DETAILS

Experimental Setup

In this research, we implemented six supervised ML algorithms to evaluate the possibility of cardiac arrest and the performance of a classifier in a classification problem. For evaluation, the real-time dataset was collected using sensors and equipment as described in section Methodology. The dataset consists of 10 attributes, among which five main attributes, i.e., the vital signs of the patient, are mainly considered. For experimental execution, the "Data Streamer" tool was used in this study to collect the pulse rate every 2 min. Data Streamer is an open-source tool used to collect live data supporting capturing, analyzing, and visualizing real-time sensor data in Excel.

As described in section Methodology, the next step uses an ML model: ANN, RFC, XGBoost, SVM, NB, and DT classifier. Before applying the model to our collected data, the collected dataset was imported using "read_csv()," and then the data was pre-processed as discussed in section Methodology. In the next step, we did feature selection using "StandardScaler()" where the vital signs of a patient were majorly considered. Furthermore, the dataset was

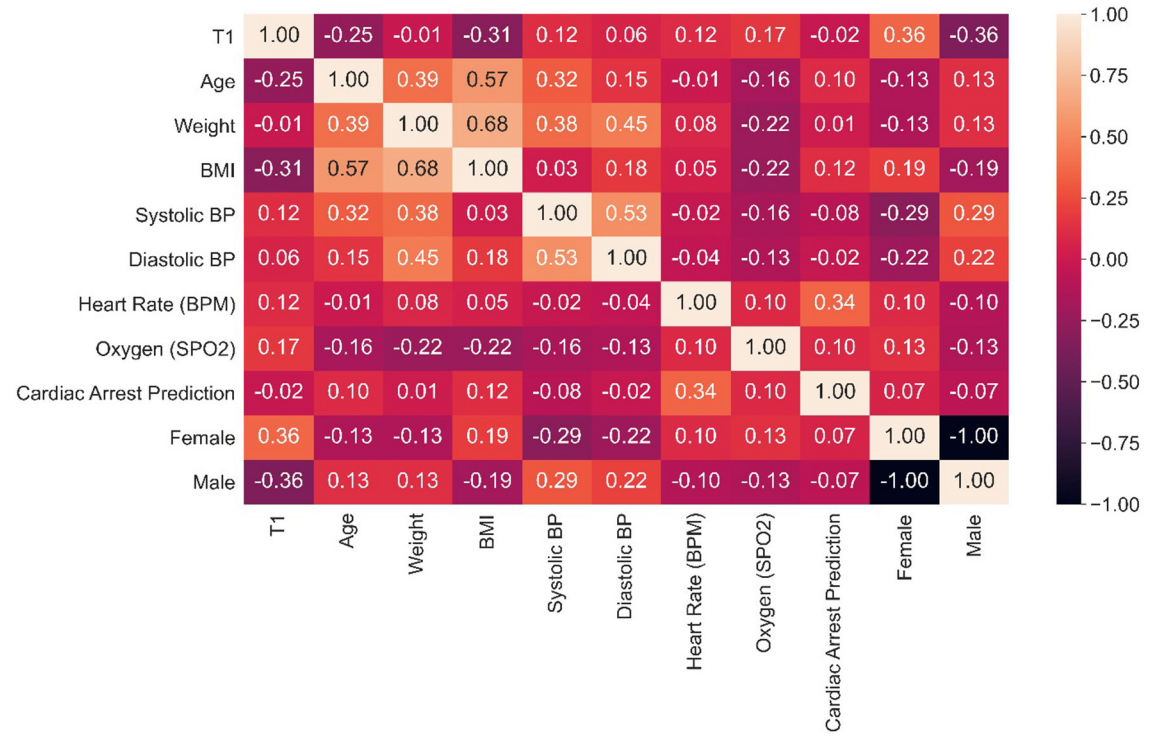


FIGURE 6 | Men/women survival probability graph.

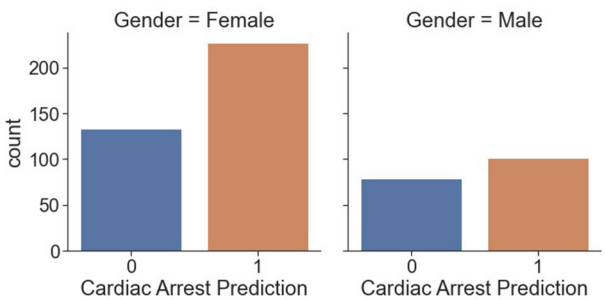


FIGURE 7 | Counts of men and women.

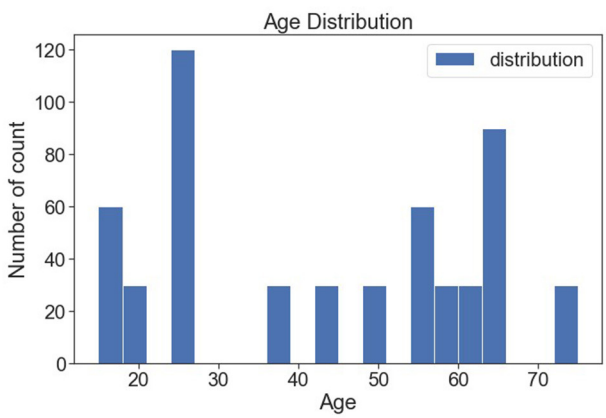


FIGURE 8 | Attributes correlation matrix.

split up into 80% training and 20% testing set. Then, models were applied to these data to check model accuracy. In the next phase, to test the performance of the classifier, various experimental metrics were computed. The contribution part of our research involves gender-wise analysis of risk/survival probability using the Kaplan-Meier algorithm, and age-wise risk probability for 10 years was formulated using the "XLSTAT" add-in tool in Excel. XLSTAT is an open-source software used to analyze, customize, and display results in Excel. Hence, the experiments discussed above have been carried out using Anaconda software in a Jupyter notebook python environment using ML libraries on an Intel(R) Core (TM) i5-3320M CPU @ 2.60 GHz system.

Evaluation Metrics

The evaluation metrics of the model are formulated with the confusion matrix. For this, the calculation of values is measured based on:

- True positive (TP) = number of occurrences that are correctly determined.
- False negative (FN) = number of occurrences that are incorrectly predicted and not required.

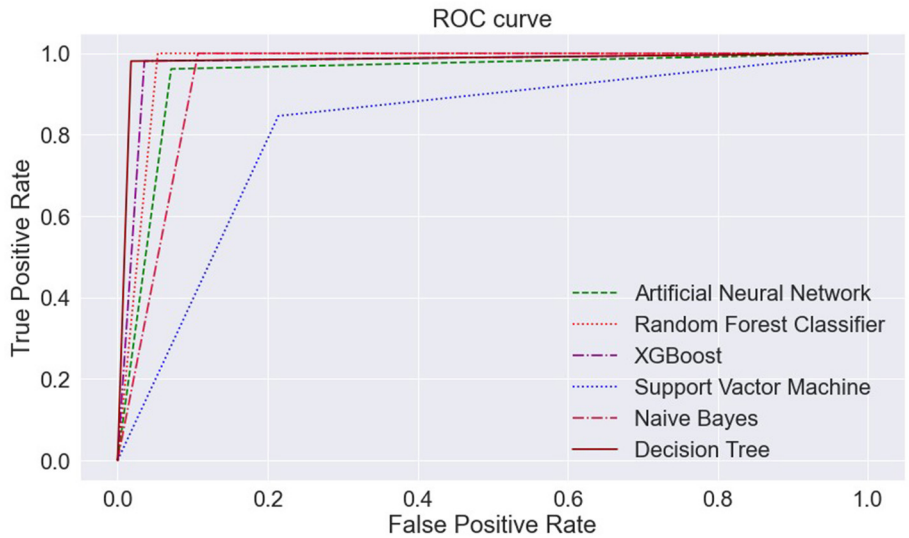


FIGURE 9 | Counts of cardiac arrest in men/women.

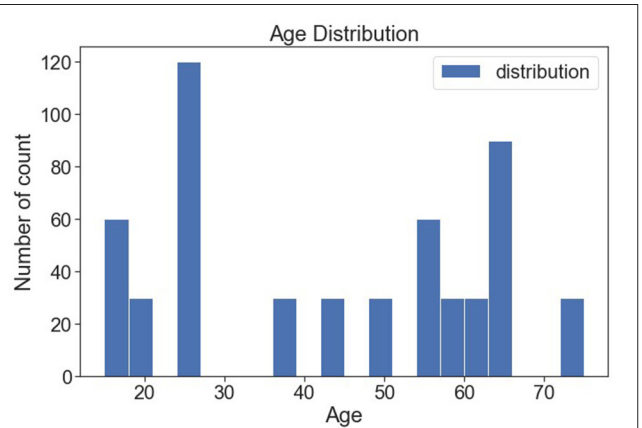


FIGURE 10 | Age-wise distribution.

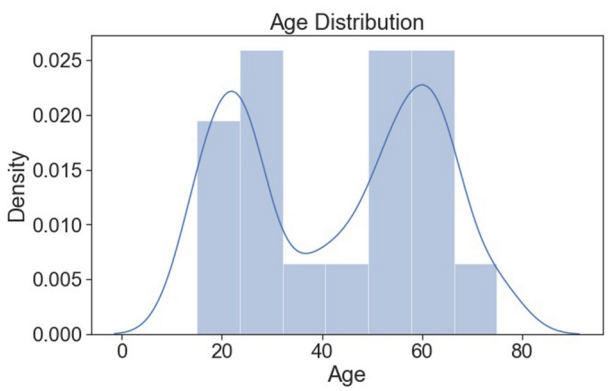


FIGURE 11 | Age distribution (normalized).

- False positive (FP) = number of occurrences that are incorrectly predicted.
- True negative (TN) = number of occurrences that are correctly predicted and not required.

Based on this parameter, we have calculated four measurements with the given formula:

- Accuracy = $TN + TP/TN + TP + FN + FP$
- Precision = $TP/TP + FP$
- Sensitivity (Recall) = $TP/TP + FN$
- F1 Score = $2^*(Precision^*Recall/Precision + Recall)$

These measurements help to know the cardiac risk associated with heart diseases.

Results

After executing the ML model on the collected dataset, we found the accuracy, sensitivity, precision, and F1-score of each algorithm, as shown in **Table 14**. These measurements were calculated with the support of confusion matrix value TP, TN, FP, and FN for each algorithm shown in **Figures 12A–F**.

Table 17 shows the Precision, Recall, F1- score, and overall accuracy of ML algorithms. The value for Precision, Recall, and F1- score is calculated against 2—the class problem of having cardiac arrest risk and no cardiac arrest risk. The label "0" has been used for no risk in cardiac diseases, and "1" has been used for risk in cardiac disease.

Hence, **Table 17** concludes that the DT outperforms all ML algorithms achieving overall 96% accuracy.

Figures 13A–F portrays the area under the curve (AUC) and receiver operating characteristic (ROC) curve. ROC curve depicts

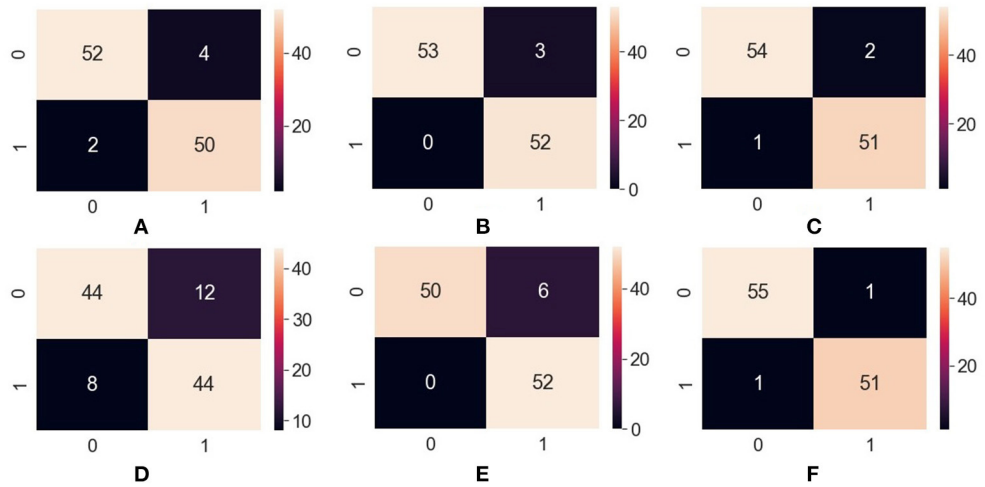


FIGURE 12 | Confusion matrix—**(A)** ANN, **(B)** RFC, **(C)** XGBoost, **(D)** SVM, **(E)** Naïve Bayes, **(F)** Decision Tree. ANN, Artificial Neural Network; RFC, Random Forest Classifier; XGBoost, Extreme Gradient Boosting; SVM, Support Vector Machine.

TABLE 17 | Model performance.

Algorithm	Overall accuracy	Label	Precision	Recall	F1-score
Decision tree	0.98	0	0.98	0.98	0.98
		1	0.98	0.98	0.98
Gradient boosting	0.97	0	0.94	0.98	0.97
		1	0.98	0.96	0.97
Random forest classifier	0.97	0	0.94	1.00	0.97
		1	1.00	0.94	0.97
Artificial neural network	0.94	0	0.92	0.96	0.94
		1	0.96	0.92	0.94
Support vector machine	0.81	0	0.78	0.84	0.81
		1	0.84	0.78	0.81
Naïve bayes	0.94	0	0.89	1.00	0.94
		1	1.00	0.89	0.94

the graphical appearance to measure the model performance. It shows the true negative, and true positive values plotted at distinct levels of the threshold.

The AUC curve represents the aggregate measures of performance at different threshold values. It also describes the probability value from the classification model to get an accurate result. If the AUC curve value shows 0 value, it predicts wrong results, and if 1, it means an exact model. Here, in **Figure 14**, the accuracy of all three classification models reaches 1 value, which means our prediction accuracy is highly accurate.

DISCUSSION, CONCLUSION, AND FUTURE WORK

Discussion

This study is aimed to detect and predict whether the patient is prone to cardiac arrest or not. This research was carried out using six ML classification algorithms, namely, ANN, RFC, XGBoost, SVM, NB, and DT classifier on a real-time collected dataset. Different tools were conducted for analyzing and classifying data on Intel(R) Core (TM) i5-3320M CPU @ 2.60 GHz system.

Dataset was split up into 0.8% training and 0.2% testing set, and then data pre-processing was done. To get the result, ML algorithms were applied to datasets to get the accuracy of the result. The results—accuracy, Precision, Recall, and F1-score—were generated using Python programming on Jupyter notebook.

Figures 15A,B show the scores. Precision, Recall, and F1-scores of the patients having no risk or chances of cardiac arrest and patients with a risk of cardiac arrest. These scores help us to understand the measure of relevance of accuracy.

Comparing both charts, we conclude that "DT" reaches the maximum accuracy based on the scores retrieved from Precision, Recall, and F1 scores.

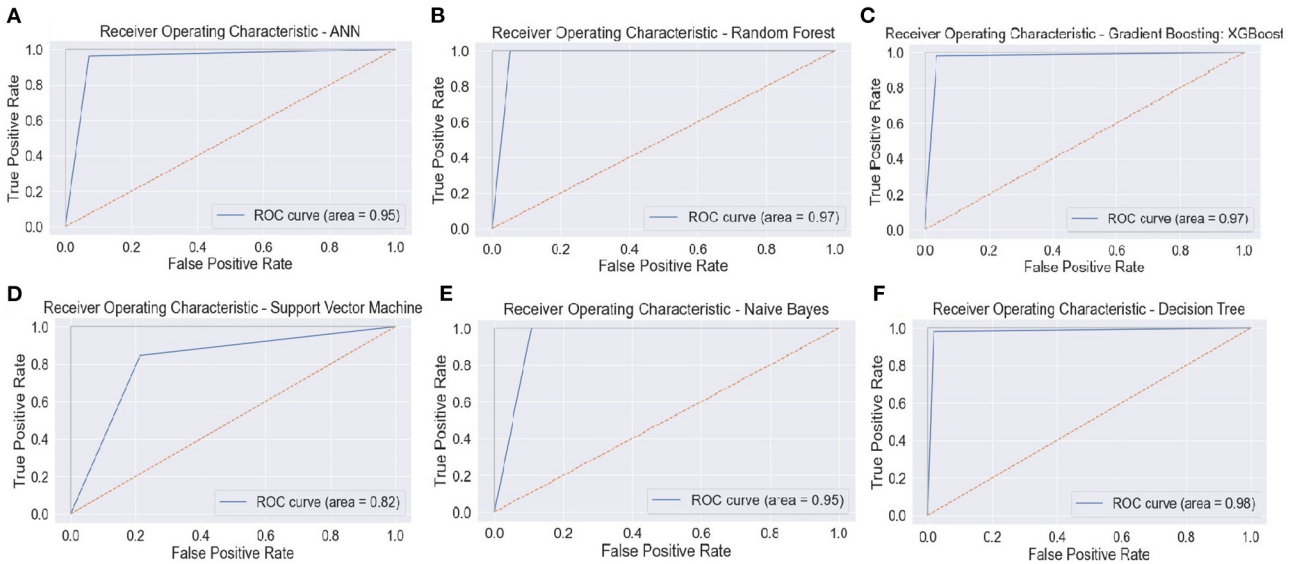


FIGURE 13 | ROC curve—**(A)** ANN, **(B)** RFC, **(C)** XGBoost, **(D)** SVM, **(E)** Naïve Bayes, and **(F)** Decision Tree. ANN, Artificial Neural Network; RFC, Random Forest Classifier; XGBoost, Extreme Gradient Boosting; SVM, Support Vector Machine.

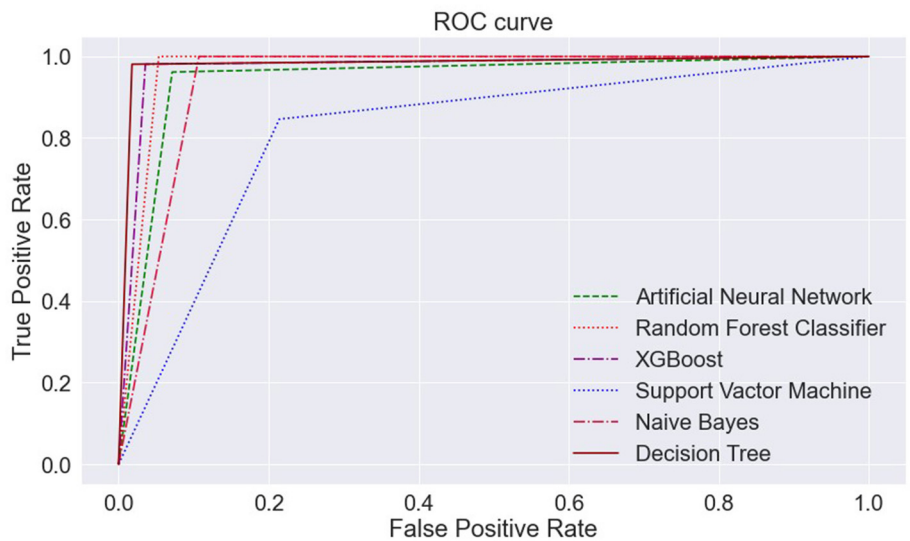


FIGURE 14 | ML models—ROC curve. ML, machine learning; ROC, receiver operating characteristic.

Furthermore, the main contribution of our research work shows an effective and efficient method for the prediction of cardiac arrest risk in gender-wise and different age-wise people in terms of survival probability. The method Kaplan-Meier shows statistical analysis in gender for showing the chances of survival probability. The result obtained by applying the Kaplan-Meier model to our collected dataset shows that female survival probability is more than that in men, as shown in **Figure 6**. Another method, Cox regression analysis, is another statistical analysis method that shows prediction/survival probability for the next 10 years of risk in cardiac arrest for different age groups. We can use this model to see cardiac risk in an individual. In addition, this method helps identify whether a person is at risk in the coming next 10 years using the score sheet displayed above.

Conclusion

This study contributes classification techniques for detecting and comparing gender-based and age-based probability for cardiac arrest survival. The main objective of our work is to

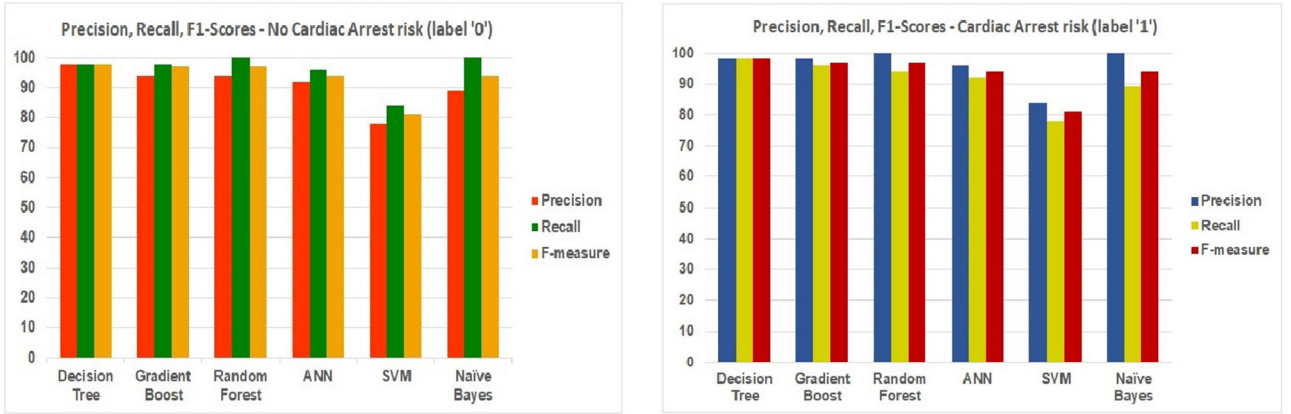


FIGURE 15 | ML models—Precision, Recall, and F1-scores—for **(A)** label "0" **(B)** label "1."

recognize cardiac arrest in a patient as early as possible using an ML model. Apart from that, our main contribution shows survival probability in an individual with gender-based and age-based using two effective methods: Kaplan-Meier and Cox regression methods. From the result, in detecting cardiac arrest risk in gender-based survival probability, female patients of our collected record show higher chances of survival than male patients. While in detecting cardiac arrest risk in age-based survival probability, patients with age 30 and more may have chances of cardiac risk as per our proposed model. From our classification model, the predicting results of the DT outperform the other ML classifiers.

Future Study

This study can be further extended by adding more diseases and making predictions using different classifier models. In addition to that, we can add more features, and predictions can be made on a larger dataset.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SP and SA: conceptualization and design. SP, SA, and AS: data collection, manuscript writing, conceptualization, and design. SP, SA, AS, and KK: manuscript proofreading, manuscript writing, conceptualization, and design. SR: revisions, conceptualization, manuscript writing, and manuscript proofreading. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Ambekar S, Phalnikar R. Disease risk prediction by using convolutional neural network. In: *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. Pune (2018). p. 1–5. doi: 10.1109/ICCUBEA.2018.8697423
2. Ramprakash P, Sarumathi R, Mowriya R, Nithyavishnupriya S. Heart disease prediction using deep neural network. In: *2020 International Conference on Inventive Computation Technologies (ICICT)*. Coimbatore (2020). p. 666–70. doi: 10.1109/ICICT48043.2020.9112443
3. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. *SN Comput Sci.* (2020) 1:345. doi: 10.1007/s42979-020-00365-y
4. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput.* (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
5. Maini E, Venkateswarlu B, Gupta A. Applying machine learning algorithms to develop a universal cardiovascular disease prediction system. In: *International Conference on Intelligent Data Communication Technologies and Internet of Things.* (2018). p. 627–632. doi: 10.1007/978-3-030-03146-6_69
6. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access.* (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707
7. Fitriyani NL, Syafrudin M, Alfian G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. *IEEE Access.* (2020) 8:133034–50. doi: 10.1109/ACCESS.2020.3010511
8. Sarmah SS. An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. *IEEE Access.* (2020) 8:135784–97. doi: 10.1109/ACCESS.2020.3007561
9. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors. *J Real Time Image Proc.* (2021) 18:1435–9. doi: 10.1007/s11554-021-01122-x
10. Chauhan U, Kumar V, Chauhan V, Tiwary S, Kumar A. Cardiac arrest prediction using machine learning algorithms. In: *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT).* (2019). p. 886–90. doi: 10.1109/ICICICT46008.2019.8993296
11. Reddy GT, Reddy MP, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart

disease diagnosis. *Evol Intell.* (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
12. Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In: *2020 International Conference on Electrical and Electronics Engineering (ICE3)*. (2020). p. 452–7. doi: 10.1109/ICE348803.2020.9122958
13. Li JP, Haq AU, Din SU, Khan J, Khan A, Saboor A. Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access.* (2020) 8:107562–82. doi: 10.1109/ACCESS.2020.3001149
14. Pan Y, Fu M, Cheng B, Tao X, Guo J. Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform. *IEEE Access.* (2020) 8:189503–12. doi: 10.1109/ACCESS.2020.3026214
15. Garg A, Sharma B, Khan R. Heart disease prediction using machine learning techniques. *IOP Conf Ser.* (2021) 1022:012046. doi: 10.1088/1757-899X/1022/1/012046
16. Yadav KK, Sharma A, Badholia A. Heart disease prediction using machine learning techniques. *Inform Technol Industry.* (2021) 9:207–14. doi: 10.17762/itii.v9i1.120
17. Latha CB, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inform Med Unlocked.* (2019) 16:100203. doi: 10.1016/j.imu.2019.100203
18. Singh H, Navaneeth NV, Pillai GN. Multisurface proximal SVM based decision trees for heart disease classification. In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. (2019). p. 13–18. doi: 10.1109/TENCON.2019.8929618
19. Sowmiya C, Sumitra P. A hybrid approach for mortality prediction for heart patients using ACO-HKNN. *J Ambient Intell Hum Comput.* (2021) 12:5405–12. doi: 10.1007/s12652-020-02027-6
20. Alotaibi FS. Implementation of machine learning model to predict heart failure disease. *Int J Adv Comput Sci Appl.* (2019) 10:261–8. doi: 10.14569/IJACSA.2019.0100637
21. Nikookar E, Naderi E. Hybrid ensemble framework for heart disease detection and prediction. *Int J Adv Comput Sci Appl.* (2018) 9:1–8. doi: 10.14569/IJACSA.2018.090533
22. Tama BA, Im S, Lee S. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *BioMed Res Int.* (2020) 2020:9816142. doi: 10.1155/2020/9816142
23. Tate AR, Rao GH. Activity trackers, wearables, noninvasive technologies for early detection, and management of cardiometabolic risks. *Int J Biomed.* (2020) 189–97. doi: 10.21103/Article10(3)_RA2
24. Prabhu T, Darshana J, Dharani Kumar M, Hansaa Nazreen M. Health risk prediction by machine learning over data analytics. *Int Res J Eng Technol.* (2019) 6:606–11. doi: 10.21203/rs.2.24678/v1
25. Shankar V, Kumar V, Devagade U, Karanth V, Rohitaksha K. Heart disease prediction using CNN algorithm. *SN Comput Sci.* (2020) 1:170. doi: 10.1007/s42979-020-0097-6
26. Amin P, Anikireddypally NR, Khurana S, Vadakkemadathil S, Wu W. Personalized health monitoring using predictive analytics. In" *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. Las Vegas, NV (2019). p. 271–8. doi: 10.1109/BigDataService.2019.00048
27. Chen Y, Zhang L, Qin X, Yi B. Preoperative risk prediction of heart failure with numerical and textual attributes. *Int J Innov Comput Inform Control.* (2020) 16:2035–46. doi: 10.24507/ijicic.16.06.2035
28. Cohen S, Liu A, Wang F, Guo L, Brophy JM, Abrahamowicz M, et al. Risk prediction models for heart failure admissions in adults with congenital heart disease. *Int J Cardiol.* (2021) 322:149–57. doi: 10.1016/j.ijcard.2020.08.039
29. Chang D, Chang D, Pourhomayoun M. Risk prediction of critical vital signs for icu patients using recurrent neural network. In: *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. (2019). p. 1003–6. doi: 10.1109/CSCI49370.2019.00191
30. Youssef Ali Amer A, Wouters F, Vranken J, de Korte-de Boer D, Smit-Fun V, Duflot P, et al. Vital signs prediction and early warning score calculation based on continuous monitoring of hospitalised patients using wearable technology. *Sensors*. (2020) 2020:6593. doi: 10.3390/s20226593
31. Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolau J, Wohlfahrt P, Wegerich S, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the LINK-HF multicenter study. *Circulation.* (2020) 13:e006513. doi: 10.1161/CIRCHEARTFAILURE.119.006513
32. Jia X, Baig MM, Mirza F, GholamHosseini H. A cox-based risk prediction model for early detection of cardiovascular disease: identification of key risk factors for the development of a 10-Year CVD risk prediction. *Adv Prevent Med.* (2019) 2019:8392348. doi: 10.1155/2019/8392348
33. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep.* (2020) 10:5245. doi: 10.1038/s41598-020-62133-5
34. Made F, Nonterah EA, Tlotleng N, Ntlebi V, Naicker N. Ten-year risk of fatal cardiovascular disease and its association with metabolic risk factors among waste pickers in South Africa. *BMC Cardiovasc Disord.* (2021) 21:336. doi: 10.21203/rs.3.rs-236769/v1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Shah, Ahirrao, Pandya, Kotecha and Rathod. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fcomp.2021.775368

Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic

Ram Krishn Mishra[1], Siddhaling Urolagin[1], J. Angel Arul Jothi[1], Ashwin Sanjay Neogi[1] and Nishad Nawaz[2]*

[1]Department of Computer Science, BITS Pilani, Dubai Campus, Dubai, United Arab Emirates, [2]Department of Business Management, College of Business Administration, Kingdom University, Ar Rifā, Bahrain

The Covid-19 pandemic has disrupted the world economy and significantly influenced the tourism industry. Millions of people have shared their emotions, views, facts, and circumstances on numerous social media platforms, which has resulted in a massive flow of information. The high-density social media data has drawn many researchers to extract valuable information and understand the user's emotions during the pandemic time. The research looks at the data collected from the micro-blogging site Twitter for the tourism sector, emphasizing sub-domains hospitality and healthcare. The sentiment of approximately 20,000 tweets have been calculated using Valence Aware Dictionary for Sentiment Reasoning (VADER) model. Furthermore, topic modeling was used to reveal certain hidden themes and determine the narrative and direction of the topics related to tourism healthcare, and hospitality. Topic modeling also helped us to identify inter-cluster similar terms and analyzing the flow of information from a group of a similar opinion. Finally, a cutting-edge deep learning classification model was used with different epoch sizes of the dataset to anticipate and classify the people's feelings. The deep learning model has been tested with multiple parameters such as training set accuracy, test set accuracy, validation loss, validation accuracy, etc., and resulted in more than a 90% in training set accuracy tourism hospitality and healthcare reported 80.9 and 78.7% respectively on test set accuracy.

Keywords: social media tourism, text analysis, deep learning, topic modeling, sentiment analysis

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Usman Mohammed,
King Khalid University, Saudi Arabia
Amir Ahmad,
United Arab Emirates University, United Arab Emirates
Amit Kumar Mishra,
DIT University, India

****Correspondence:***
Ram Krishn Mishra
rkmishra@dubai.bits-pilani.ac.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Computer Science

Received: *13 September 2021*
Accepted: *19 October 2021*
Published: *05 November 2021*

Citation:
Mishra RK, Urolagin S, Jothi JA, Neogi AS and Nawaz N (2021) Deep Learning-based Sentiment Analysis and Topic Modeling on Tourism During Covid-19 Pandemic. Front. Comput. Sci. 3:775368. doi: 10.3389/fcomp.2021.775368

INTRODUCTION

The tourism sector globally plays a significant role in long-term economic development. A huge movement of tourists from one place to another for leisure, business, family, or friend has connected the globe as a single village. Tourists like to visit the places for their happiness or business. Tourism is held accountable for many factors, such as it increases income, and helps in the development of many other industries (Manzoor et al., 2019). The advantages of travel and tourism extend well beyond their direct effects on GDP and employment; they also include indirect benefits through supply chain links to other industries and create induced effects. As per the report (Poole & Misrahi, 2020), the tourism industry contributed 10.3% to global GDP, and domestic travel accounts for the bulk of worldwide Travel and Tourism spending in 2019, accounting for 71.3% of total spending globally, with foreign tourists accounting for the remaining 28.7%. Many countries focus on foreign tourism because of its cash generation. Domestic tourism may be a vital instrument for generating employment, economic growth, poverty reduction, and infrastructure upgrades.

The Covid-19 has caused the globe to face a significant and complex dilemma (Aburumman, 2020). It has turned out to be a health issue, and it has also ruined the entire corporate process. Travel and tourism have been severely impacted in the corporate world (Alamanda et al., 2019). Due to the total lockdown enforcement by every country, every stakeholder in the tourist sector cannot revive. Tourism is a significant and growing source of revenue in the economy. According to UNCTAD research released on June 30, 2021, a drop in international tourism caused by the coronavirus pandemic may cost the world economy more than $4 trillion in 2020 and 2021 (*COVID-19 and Tourism - An Update*, 2021). Prior to the Covid-19 pandemic, travel and tourism had grown to be one of the most significant sectors in the global economy, accounting for 10% of global GDP. Only 25 million individuals traveled abroad in 1950, at the dawn of the jet era. By 2019, that figure had risen to 1.5 billion, and the travel and tourism industry had grown to nearly unfathomable dimensions for many nations globally.

Among the G20 nations, the hospitality and travel industries account for 10% of employment and 9.5% of GDP on average, with Italy, Mexico, and Spain accounting for 14% or more of GDP. A 6-month disruption inactivity may directly lower GDP by 2.5–3.5% (*IMF F and D.,2021*).

The hospitality sector has been put under tremendous strain due to the exponential decline in personal and business travel, resulting in reduced occupancy and revenue per available room. During the lockdown phase, 98% of global international tourists had declined in May 2020 as compared to 2019 which leads to a big loss in hospitality (UNWTO., 2020). This equates to a drop of 300 million tourists and a loss of 320 billion dollars in international tourism revenues, which is more than three times the loss experienced during the global economic crisis of 2009. Oversupply, declining hotel prices, low consumer sentiment, travel restrictions, and the delay of key global events are all causing severe operational and financial difficulties for industry players. As a result, In Malaysia, hotel owners have been compelled to cut expenses by taking actions such as reducing staff numbers and wages, forcing workers to work without pay, and closing hotels entirely (Foo et al., 2020).

Passengers' behavior has shifted as a result of the Covid-19 problem, travel limitations, and the ensuing economic crisis, resulting in a significant decrease in demand for airline services. As measured by revenue per passenger kilometers, passenger air transport was down 90 percent year on year in april 2020, and it was still down 75 percent in August, according to IATA (IATA, 2021.). The decline in economic activity and trade had an effect on freight, the freight service was almost 30% lower year on year in april and was still around 12% lower in August. To fight against the spread of the coronavirus, travel restrictions have been imposed in all nations. Airport closures, suspension of incoming and outgoing flights, and nationwide lockdowns are just a few of the measures governments are doing to combat the epidemic. Following the pandemic's spread in the first two quarters of 2020, at least 93 percent of the global population lived in countries with coronavirus-related travel restrictions, with almost three billion people living in governments that imposed entire border bans on foreigners.

According to the research jointly published by UNCTAD, 2021a and the UN World Tourism Organization (UNWTO), international tourism and its closely related industries incurred a 2.4 trillion dollars loss in 2020. The tourism sector may face a similar loss this year, and that the tourist industry's recovery would be dependent on the global uptake of coronavirus vaccinations (Global Economy Could Lose over $4 Trillion Due to COVID-19 Impact on Tourism | UNCTAD,2021b, n. d.-a). Due to the mounting losses in many countries, Covid-19 vaccines are more prevalent in certain countries than others, tourist losses in most industrialized countries are minimized but aggravated in underdeveloped countries. According to the research, the tourist industry is expected to recover faster in nations with high vaccination rates, including the United Kingdom, the United States, France, Switzerland, and Germany. However, many experts predict that recovery to Pre-Covid international visitor arrival levels would not occur until 2023 or later, according to the UNWTO ("UNWTO World Tourism Barometer and Statistical Annex, December 2020," 2020).

The study examines the perceptions of Twitter users in the Covid-19 pandemic concerning two tourism sub-domains: hospitality and healthcare. The research was carried out using deep learning algorithms and Natural Language Processing (NLP) methods. The research objectives are to identify the common similar shared terms related to tourism sub-domain healthcare and hospitality using a topic modeling. We find the most significant themes in each sector and reveal hidden semantic patterns inside both sectors. This will aid us in locating and comprehending the most widely discussed issues in the area. This research also looks at the direction of people's feelings towards tourism healthcare and hospitality during the covid-19 outbreak by calculating the sentiment score and classifying in the different emotional categories. Now that it has been trained on real-world data, this model may be utilized in prediction of user's emotions on new tweets. Lastly, we will talk about the study's challenges and issues and the contributions of this research in possible future works. The structure of the research is organized in a systematic flow, **section 2** describes the literature review, **section 3** methodology, **section 4** experimental result, **section 5** discussion and at the last conclusion and future scope.

LITERATURE REVIEW

The literature review discusses the domains related to the research study. It also provides a panoramic overview of the theoretical studies done in the tourism subdomain. Some relevant studies which are related to this research study's methodology have been pondered upon. Along with that, the theoretical studies have been discussed accordingly. It provides a rich source of information. Therefore, an extensive review of literature has been done in the area of social media, sentiment analysis, topic modeling, deep learning, and tourism during the pandemic crisis.

The highly contagious coronavirus continues to wreak havoc on the tourism sector, raising serious questions about its present and long-term survival. The infection has resulted in the loss of millions of jobs. India has a hospitality industry which is one of the largest

tourism industries of the world and provides employment to lakhs of people. Ever since the Covid-19 started, India's hospitality and the airline industry suffered a massive loss in revenue and has resulted in a 30 percent dip in tourists. Being a top tourism destination, Malaysia has suffered a net loss of 5.5 million USD in hotel cancellations alone due to growing concern about the raging virus (Foo et al., 2020). Kuala Lumpur is one of the top tourist destinations for many people around the world. On top of that, the losses incurred are far greater when the government of Malaysia shut its doors by banning international flights. A study was conducted on a cross-regional comparison of the effects of covid-19 on the tourist sector. The authors of the study concluded that the tourist industry is vulnerable to global crises (Uğur & Akbıyık, 2020). With the dissemination of the news, passengers opt to postpone or delay their journeys on the same day. Comments on the benefits of travel insurance and refunds due to trip cancellations were among the issues that further in-depth studies discovered. An algorithm for sentiment analysis and fake news classifications has been proposed in the studies (Hakak et al., 2021). Another study provided a random forest approach for predicting the patient health during Covid-19 (Iwendi et al., 2020) (Khan et al., 2020). The Deep learning approaches provide to deal with sensitive data specifically healthcare such as early detection of diabetic (Gadekallu et al., 2020) where prediction and model accuracy should be more accurate. Once the model is ready, optimization help to find the ideal iteration and finalizing the model with proper perfections (Agrawal et al., 2021).

With regards to social media and then amount of data it generates every day, twitter users are said to create around 15 GB of data each day. It is widely utilized by the general people, who use it to express their opinions (Pandey et al., 2019)on a variety of public topics (Srivastav et al., 2020) and to voice their complaints to businesses and government agencies. Since tourism is one of the biggest industries in terms of generating revenue towards a country's economy and at the same time most of the data is generated from various social media websites. A study (Alaei et al., 2017) has employed various sentiment analysis techniques in tourism, which are examined and evaluated in terms of datasets used and performance on key assessment measures. Facebook has been witnessed to be used by the Korean government for the promotion and growth of tourism (Hoon et al., 2016). Due to the Coronavirus pandemic, a huge transformation towards online usage has been witnessed along with a sudden increase in Twitter usage. Social media is a rich source of information for tourists planning for any travel strategy. Machine learning and deep learning have played an important role in the field of medical image processing where machine learning algorithms have been used to classify Covid-19 affected individuals (Bhattacharya et al., 2021) and classifying the severity of Covid-19 diagnosed patient using neuro fuzzy techniques (Ayoub et al., 2021; Iwendi et al., 2021). Another study on how deep learning is being used for healthcare delivery in low and middle income countries has been done where the authors have explained how pneumonia detection using state-of-the-art deep learning algorithms can be utilized to help those countries with limited facilities and resources (Williams et al., 2021).

A study in (Kaushal & Srivastava, 2021) has taken a rather interesting approach to study the impact of Covid-19 on the tourism industry by taking on the interviews with 15 participants in senior positions in the hospitality industry. The study takes a qualitative approach to research, with email interviews serving as the primary method of data gathering. Email interviews are becoming more frequent in qualitative research, and given the present climate of social distance, it is deemed most suited for the current inquiry. Email interviews are also considered to be less expensive than telephonic or face-to-face interviews, yet they can generate detailed information from participants. We all are aware of the fact that the coronavirus indeed emerged in the sprawling capital of Central China's Hubei province–Wuhan. The study aimed to assess the impact of coronavirus outbreaks on the Chinese tourism sector (Yang et al., 2020). Global visitors have abandoned their plans to visit China, while Chinese tourists are barred from visiting other countries. The fast spread of the Coronavirus in China has put a halt to people's daily lives. This concern has influenced the country's tourism sector both locally and abroad. According to the conclusions of the study, because of the danger from the virus, outsiders are afraid to interact with the Chinese population.

Countless individuals have lost their jobs as a result of the epidemic. Businesses are facing difficulties in returning to pre-covid levels, and they are looking for new revenue-generating techniques. While looking for new methods, it is critical to pay attention to the conversations taking place on social media to grasp the emotion. This may give suggestions for the development of new revenue-generating initiatives for firms. In a study (Jeong et al., 2019) on the topic modeling and sentiment analysis of social media data, the authors of this paper presented an opportunity mining technique for identifying product possibilities. An opportunity algorithm based on the significance and satisfaction of product subjects identifies the opportunity value and improvement direction of each product topic from a customer-centered perspective. We must be ready to adapt to new conditions in these unusual times when a whole country's economy has stalled due to a lack of tourism revenue. Businesses, too, should adapt by adopting the aforementioned strategy to develop new ways and items to market for increasing their revenue stream and attracting tourists.

METHODOLOGY

Many techniques were utilized in the research to delve deeper and gain a better understanding of the Covid-19 epidemic's impact on high-level tourist firms, including topic modeling, sentiment analysis, and deep learning classification. The authors have divided the tourist data into four categories: hospitality and healthcare. **Figure 1** illustrates a high-level overview of the many techniques taken by the authors to understand the impact of the tourist industry.

The datasets for each sector have been saved in a CSV format with four attributes: datetime, tweet id, text, and username.

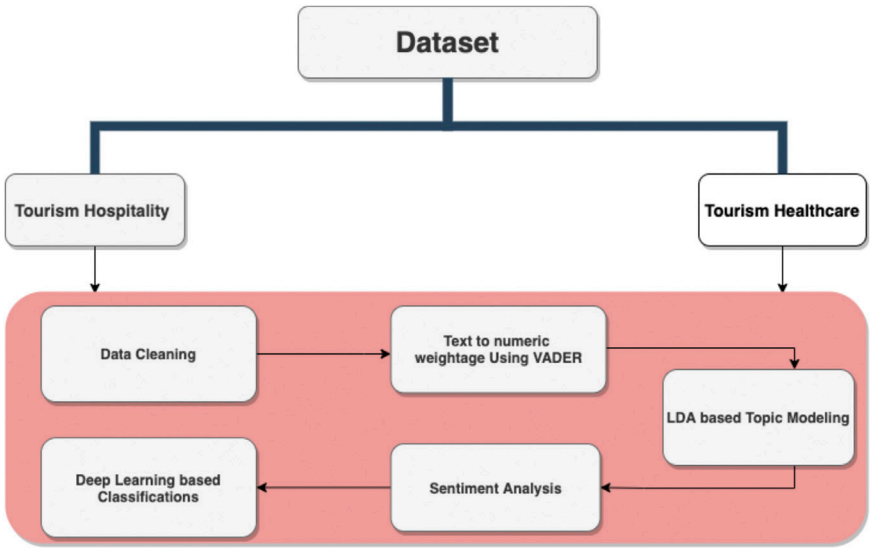


FIGURE 1 | Processing structure.

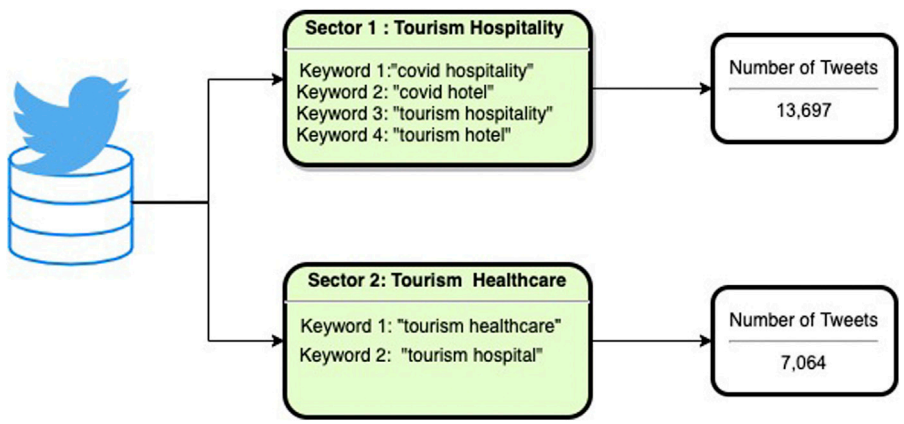


FIGURE 2 | Data collection.

Data Collection

Over the course of 8 months, we have gathered a total of 20,761 tweets pertaining to tourism, healthcare and hospitality. The data was gathered with the help of the Tweepy library, an open-source Python program that connects to the Twitter database. The timeline of the tweets was between april 2020 and December 2020, when the Covid-19 virus began to spread all around the world, forcing governments to seal their borders and ensure travel restrictions. **Figure 2** depicts the data collecting strategy. Since the main purpose is to understand how the tourism sector has been affected by the coronavirus, it has been segmented into two parts.

The tourism hospitality business is the first segment, with 13,697 tweets. The tourism healthcare services is the second segment that is helping the people during pandemic time. We found 7,064 tweets about healthcare using the keywords' tourism healthcare', and "tourism hospitals". We focused on collecting these two sectors to identify the interdependencies between the similar kinds of topics shared in twitter.

Data Preprocessing

One of the most important aspects of data analysis is ensuring that the acquired data is machine-readable. Words, photos, and videos are incomprehensible to machines; they can only understand 1 and 0s. To be able to give an input of 1 and 0s, we must go through many phases. The data must be pre-processed, which demands the use of a method known as data cleaning, which entails transforming raw data into a machine-readable format. We need to clean the big text dataset, which is made up of tweets, to remove any discrepancies and avoid having inconsistent data. The study's approach to data cleansing is straightforward. The study made use of Excel's inbuilt duplicate data removal feature and removed all complete tweet duplicates. Duplicate data can have

little to adverse effects on the machine learning model. When duplicate data was observed, the training data might have the same instances again and again. i.e. there will be multiple same values of dependent and independent variable combinations. Hence, when the proposed model learns from this data, the study will get very high accuracy on in-sample testing but out of sample testing will be much lesser, i.e., the study will have an over-fitted model.

Furthermore, we moved ahead to clean the data in much more detail by using python's open-source package called Regex which is nothing but regular expressions. The research study created a custom python script that takes in the tweets as a whole and by using regular expression (ThompsonKen, 1968). Then it started with removing special characters, while converting all the tweets to lowercase mainly because machine learning models might treat a word which is at the beginning of a tweet with a capital letter different from the same word which appears later in the same tweet but without any capital letter. This might lead to a decline in accuracy. So, therefore lowering the words would be a better trade-off. Punctuation along with leading and trailing white spaces in the tweets were also removed.

There are data cleaning steps that should be done specifically in Twitter such as removing user mention '\text [at]', removing HTTP links, removing emotions, the 'rt' character which displays when a user retweets a tweet.

NLP Techniques for Data Cleaning

As mentioned earlier, machine learning models struggle in understanding the raw text, and this is why there are a couple of natural language processing sub-techniques that we incorporated:

Tokenization

The basic components of natural language processing are tokens. Tokenization is the process pf breaking down a large chunk of text into smaller tokens. Tokens can be words, characters, or sub words.

The most frequent technique of processing raw text is at the token level, because tokens are the building blocks of Natural Language. RNN, GRU, and LSTM, among the most common deep learning architectures for NLP, analyze raw text at the token level as well. Unstructured data and natural language text are broken down into bits of information that may be regarded distinct parts using a tokenizer. The token occurrences in a document can be utilized to create a vector that represents the document. An unstructured string (text document) is instantly transformed into a numerical data structure appropriate for machine learning.

In tokenization, the authors needed to identify the words that constitute a string of characters, and it is the most basic step to proceed in processing textual data. This is important because the meaning of the text could easily be interpreted by analyzing the words present in the text (Webster & Kit, 1992). The study performed tokenization before removing stop words.

Removal of Stop Words

Stopwords are the most common words in any natural language. For the purpose of analyzing text data and building NLP models, these stopwords might not add much value to the meaning of the document. On removing stopwords, dataset size decreases and the time to train the model also decreases. A python list was created which contained all the words that fall under the category of stopwords for removing them. Removing stop words can potentially help improve the performance as there are fewer and only meaningful tokens left. Thus, it could increase classification accuracy.

Topic Modelling Process

The goal of the analytics industry is to extract "Information" from data. It's challenging to get relevant and needed information with the rising volume of data in recent years, much of which is unstructured. However, technology has created some effective ways for mining through data and retrieving the information that were needed.

Topic Modelling is one such text mining approach. As the name implies, a technique for automatically identifying themes contained in a text item and deriving hidden patterns displayed by a text corpus. As a result, wiser decisions may be made.

Topic modeling is distinct from rule-based text mining techniques that rely on regular expressions or dictionary-based keyword searches. It's an unsupervised method for spotting and tracking a group of words in huge groups of texts. Using a probabilistic model, topic modeling finds abstract themes that recur in a corpus of documents. It's commonly used as a text mining technique to uncover semantic patterns in large amounts of material. The Latent Dirichlet Allocation (LDA) statistical model was used to find themes in a collection of documents (in our example, a corpus of tweets).

Latent Semantic Structures

Latent semantic analysis (LSA) is a statistical model of word usage that allows for semantic similarity comparisons across bits of textual information. LSA's basic premise is that there is some underlying, or "latent," structure in the pattern of word usage across texts and that statistical approaches may be used to estimate this latent structure. In this scenario, documents may be regarded as contexts in which words appear, as well as smaller text pieces such as individual paragraphs or phrases.

On the basis of massive corpus studies, Latent Semantic Analysis (LSA) creates a high-dimensional vector representation. However, LSA analyzes co-occurrence across the corpus using a defined window of context (e.g., the paragraph level). The co-occurrence matrix is then subjected to a factor analytic approach (singular value decomposition) to produce a smaller set of dimensions. Words that are used in comparable contexts, even if they are not used in the same context, have similar vectors as a result of the dimension reduction. Nonetheless, their vector representations in LSA would be comparable (Landauer et al., 2009). This foundation makes it possible to compare larger chunks of text, such as the meaning of phrases, paragraphs, or whole texts.

As a theoretical model and a technique, LSA has been used to characterize the semantic relatedness of linguistic units. Its results on conventional vocabulary and subject matter exams overlap those of humans, it replicates human word sorting and category judgments, it simulates word-word and passage-word lexical priming data, and it adequately evaluates textual coherence and learnability of texts. Our methodology incorporates Latent Dirichlet Allocation (LDA), which employs the latent semantic structure for textual similarity comparison as previously mentioned.

LDA Based Topic Modelling

In this research, the authors have used the LDA model to perform topic discoveries from the existing tweets. Latent Dirichlet Allocation (LDA) is a statistical generative model using Dirichlet distributions (Jelodar et al., 2017).

The authors begin with a corpus of D documents and decide how many X topics we want to uncover from it. The topic model and D documents, represented as a mixture of the X topics, will be the outputs. LDA finds the weight of connections between documents and topics and between topics and words. The advantage of LDA is that it leverages feature information to construct a new axis, reducing variance and increasing class distance between the variables.

LDA focuses primarily on projecting the features in higher dimension space to lower dimensions. The three steps for achieving this are:

1) To begin, we must determine the separability between classes, which is defined as the distance between the mean of two or more classes. The between-class variance is the term for this is given in **Equation (1)**.

$$S_b = \sum_{i=1}^{g} N_i * (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (1)$$

2) Second, we must compute the distance between each class's mean and sample. Within-class variance is another name for it which is shown in **Equation (2)**.

$$S_w = \sum_{i-1}^{g} (N_i - 1) * S_i = \sum_{i-1}^{g} \sum_{j-1}^{N_i} \left(x_{\bar{i},j} - \bar{x}\right)\left(x_{\bar{i},j} - \bar{x}\right)^T \quad (2)$$

Where:

S_b = Seperability between the classes.
S_w = Distance between each class's mean and sample.
N = Number of instances.
$\bar{x}$= Mean of the input data points.

3) Finally, a lower-dimensional space must be created that increases variation across classes while minimizing variance within classes. P, commonly known as Fisher's criteria, is a lower-dimensional space projection.

We utilized the pyLDAvis package for our research, which is an open-source Python framework for interactive topic model visualization.

Sentiment Calculation Using VADER

Lexicon and Rule-Based Sentiment Calculation

For this work, Valence Aware Dictionary and Sentiment Reasoner (VADER) based sentiment calculation were used. VADER is a vocabulary and rule-based sentiment analysis tool that is tuned in to social media sentiments. It is an open-source utility that is completely free (Bonta et al., 2019). Word order and degree modifiers are taken into account by VADER as well. It is a text sentiment analysis model that's sensitive to both emotion polarity (positive/negative) and intensity (strong). It is included in the NLTK package, and the biggest advantage about it is that it can be used on unlabeled text data right away. The model automatically gives us a dictionary which contains the scores of each sentiment i. e negative, neutral and positive. These scores are calculated based on a metric called polarity. The range of polarity is [−1,1], with -1 indicating a negative feeling and one indicating a positive sentiment. The intensity of a word affects whether it has any effect on the next word. Adverbs are utilized as modifiers in English, such as "very good," which employs a lexicon-based method. In certain situations, the values are precisely equal to zero. A sentiment score will be assigned based on the polarity value, and the computation will be done in such a way that if the score is less than 0, the sentiment will be reported as negative. Positive sentiment is returned if the polarity is greater than 0. The score will be 0 in all other situations, and the sentiment will be neutral. There is another score returned by VADER and that is the compound score. Compound score is simply the sum of all the sentiment scores and it ranges from [-1,+1] where -1 indicates extremely negative score and on the other hand +1 indicates extremely positive score. We have used this metric primarily to understand the sentiments of the users (Hutto & Gilbert, 2014).

The main advantage of utilizing a lexicon-based approach is that it is much easier to understand and update by people. Using this method, the semantic orientation may be gathered and classified as neutral, positive, or negative. Sentiment analysis is a technique for extracting subjectivity and polarity from text, whereas semantic orientation assesses the content's polarity and strength. This approach uses adjectives and adverbs to indicate the text's semantic direction (in our case it is a tweet). The sentiment orientation value is then calculated by mixing adverbs and adjectives. Furthermore, the entire value is derived from a single source. Finally, a pooling technique is used to assess the sentiment of the tweets, which takes the average of all sentiments.

Classification of Sentiments Using LSTM RNN Algorithm

Long Short-Term Memory Networks (LSTMs) are a kind of RNN that can learn long-term dependencies. They work tremendously well on a large variety of problems and are now widely used. Neural Networks are a strong technology that may be used for image identification and a variety of other tasks. One of the model's limitations is that it has no memory connected with it. This is a difficulty while dealing with sequential data such as text or time series. RNN solves this problem by incorporating a feedback glance that acts as a memory. As a result, the model's previous inputs leave a trace. LSTM expands on this concept by including a short-term and long-term memory component (Sherstinsky, 2018).

All recurrent neural networks are made up of a series of repeated neural network modules. This repeating module in conventional RNNs will have a relatively basic structure, such as a single tanh layer. LSTMs are specifically intended to prevent the problem of long-term reliance. They do not have to work hard to remember knowledge for lengthy periods of time; it's virtually second nature to them. The entire architectural explanation of LSTM Network having different input and output layers and dealing with sequential information it combines forget gate, input gate, and output gate (Yu et al., 2019).

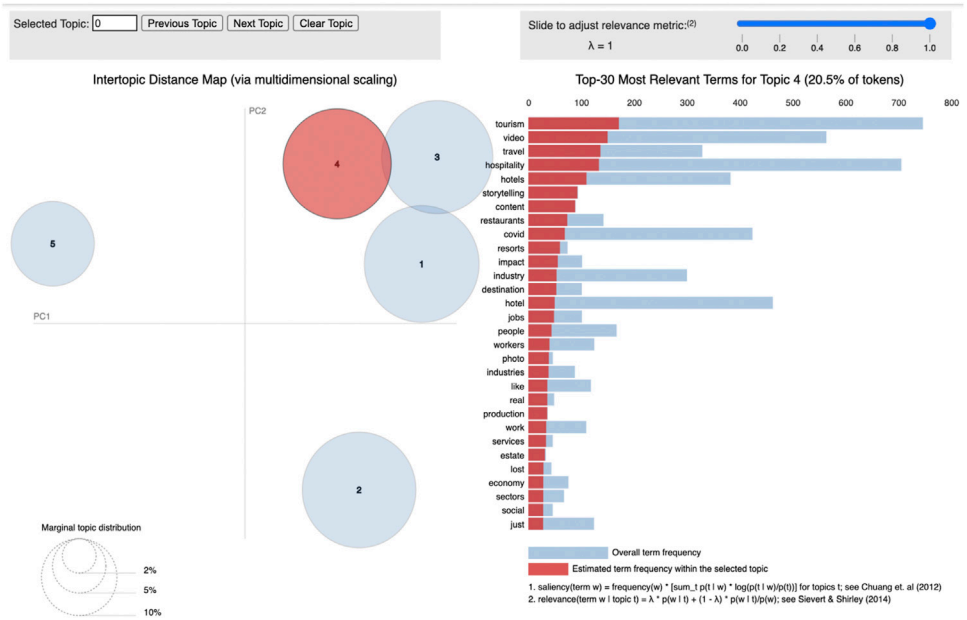


FIGURE 3 | Interactive topic modeling visualization for tourism hospitality sector.

In the case of RNN, the input is divided into two sections, and it must process multiple inputs while only performing one forward computation (Sherstinsky, 2018). This does not occur in the BP network. One of the two segregated inputs is the calculation's output, while the other is the variable present in the trained sample. RNN manages continuous data, specifically lists and sequences. This is what the characteristics mean. As a result, RNN and LSTM are particularly well-suited to dealing with the aforementioned data. However, there are a few questions to consider when using RNN. The gradients are exploding and disappearing, which is a big concern. Since the RNN is made up of several BPNNs, the calculation requires a variety of activation functions, such as sigmoid. These activation functions change the value in a certain set, such as [0, 1], and so on. Furthermore, these functions limit the spectrum of their derivatives.

LSTM has many uses and is used in many types of research, including forecasting stock markets, sentiment analysis, music generation, forecasting tourism demand, time series prediction, controlling robots, speech recognition, handwriting recognition, semantic parsing, and much more.

In some cases, we will have to examine recent data in order to complete the work at hand. Consider a vocabulary model that predicts the next word based on the ones that came before it. Recurrent Neural Network (RNN) can learn to utilize prior experience in situations when the period between historical information and the area where they're needed is brief. Because the RNN has a long-term dependency, the LSTM was included to help with this. The LSTM architecture is a kind of RNN that is capable of learning long-term relationships as well as solving the vanishing/exploding gradient problem. When working with huge volumes of data, LSTM can model and extract valuable information from social media data extremely effectively. Hence, we decided to use the LSTM RNN model since it had approximately 20,000 tweets.

Many attempts have been made to address the problem of vanishing gradients by modifying vanilla RNNs such that gradients do not expire as sequences become longer. The most popular and effective of these projects has been the long, short-term memory network, or LSTM. LSTMs have virtually supplanted regular RNNs in terms of popularity. LSTMs are also widely used to analyze social media data.

EXPERIMENTAL RESULTS AND ANALYSIS

This section discussed about the experimental results and analysis of the research study.

Depiction of Topic Modeling for Tourism Hospitality

Hovering over the first subject in **Figure 3** is the pyLDAvis interactive topic modeling visualization for the tourist hospitality industry. This figure reveals that there are 5 cluster themes connected with the sector, and the red bar represents the most common terms in each subject. The blue bar represents the overall phrase frequency of that specific word across all themes combined. Another thing to keep in mind is that if any of the clusters overlap or are physically adjacent to another cluster, it suggests that the topics contain comparable terminology.

The key themes in the first topic cluster include terms like Covid, hospitality, hotels, and industry. This implies that the Covid-19 epidemic has caused the hospitality industry to close its doors, resulting in billions of dollars in lost revenue (*Global Economy Could Lose over $4 Trillion Due to COVID-19 Impact on Tourism* | *UNCTAD.,2021c*, n. d.-b). It was observed that

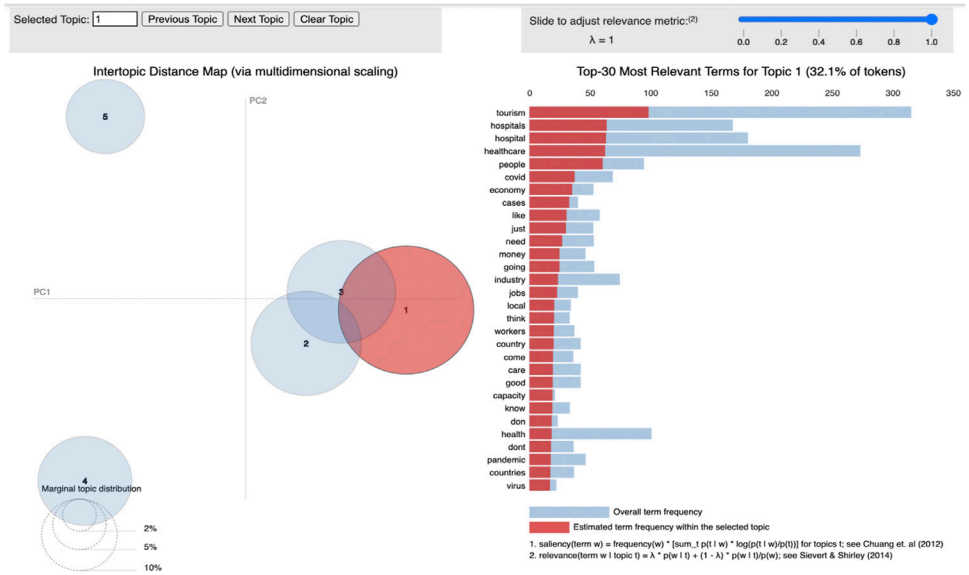


FIGURE 4 | Interactive topic modeling visualization for tourism healthcare sector.

restaurants, hotels, and resorts have been the most negatively impacted businesses owing to a variety of factors including.

1) Lack of visitors due to Covid - 19,
2) Loss of consumers (tourists) from other regions of the nation due to the pandemic.

We also discovered that employees and workers in the restaurants, hotels, and other sub-sectors were laid off in large numbers as a result of management choices made due to a shortage of finances. Another finding from the second subject cluster was that the same management implemented new methods to keep their company afloat by integrating new survival marketing tactics.

Depiction of Topic Modeling for Tourism Healthcare

Figure 4 depicts the tourism healthcare industry as an interactive graphic. The medical tourism industry was also impacted by the outbreak. By completing topic modeling, it was observed that the third topic cluster led us to Thailand, where it was learned that Thailand had made gradual steps to reopen for medical tourism, allowing restricted numbers of patients to visit the nation for treatment from the end of July 2020 (Tatum, 2020). Medical tourists and up to three others who can accompany them must follow strict standards, which include confirmation of negative Covid-19 testing granted no more than 72 h before departure, repeated tests while in-country, a 14-days hospital quarantine, and other anti-infection measures. The government is responding to the changing climate by leveraging digital technology to extend and improve telemedicine services, fostering a more hospitable business environment to encourage medical investment and innovation, and advancing its sophisticated healthcare infrastructure.

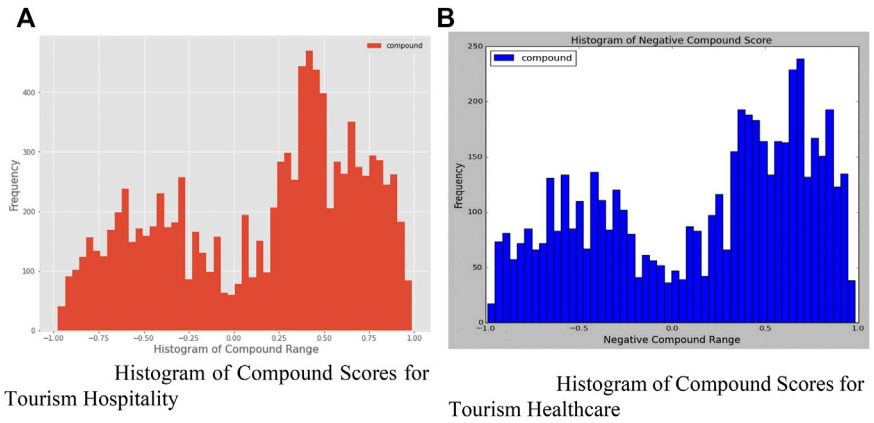
Histogram of Sentiment Compound Score for Each Sector

Figure 5 explains the compound range of sentiments. As per depiction, the compound range varies between -1 and +1. **Figure 5A** shows tourism hospitality's sentiments where positive rates are higher than negative. Tourism healthcare in **Figure 5B** has more positive ranges and the highest frequency is between 0.5 and 0.7 and most of the tweets belong to the positive side.

Sentiment bar Plot for Each Sector

Figure 6 is a bar plot depicting the number of feelings expressed in the various sectors of the industry taken into the study. The bulk of the feelings are positive, as seen in the graph. In the instance of the tourism hospitality sector in **Figure 6A**, VADER sentiment analysis revealed that about 6,300 + emotions were positive. One of the reasons for the increase in positive tweets is that when covid-19 hit the globe, there was an inflow of tweets praising the hospitality sector, particularly the hotel industry, collaborating with the government for a win-win situation. Passengers had to go through obligatory institutional quarantine, and the government was having trouble accommodating all of the quarantine passengers. The hotel sector offered the government a feasible alternative for directing travelers to quarantine themselves in hotel rooms, therefore increasing their revenue.

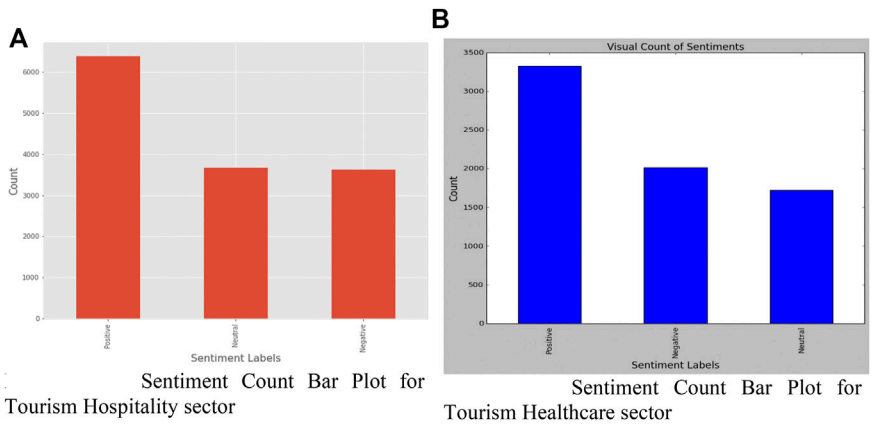
The ideal situation can be seen in the healthcare sector bar plot **Figure 6B**, where a country's GDP drop, which led to substantial system reorganization, can be linked back to a huge number of positive tweets. The negative tweets may be linked back to numerous industry owners whose businesses were harmed as a result of the lockdown, adversely impacting the country's economy. Another reason is that prominent influencers and celebrities, notably in India, began tweeting about the



Histogram of Compound Scores for Tourism Hospitality

Histogram of Compound Scores for Tourism Healthcare

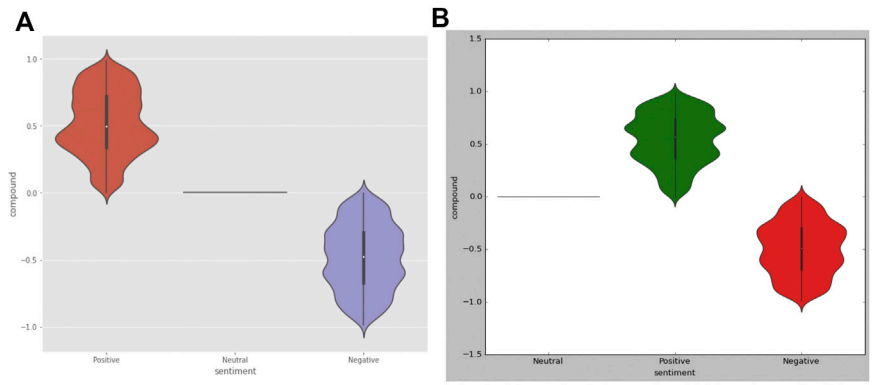
FIGURE 5 | (A) Histogram of compound scores for tourism hospitality **(B)** Histogram of compound scores for tourism healthcare.



Sentiment Count Bar Plot for Tourism Hospitality sector

Sentiment Count Bar Plot for Tourism Healthcare sector

FIGURE 6 | (A) Sentiment Count Bar Plot for Tourism Hospitality sector **(B)** Sentiment Count Bar Plot for Tourism Healthcare sector.



Compound Score Violin Plot for Tourism Hospitality Sector

Compound Score Violin Plot for Tourism Healthcare Sector

FIGURE 7 | (A) Compound score violin plot for tourism hospitality sector **(B)** Compound score violin plot for tourism healthcare sector.

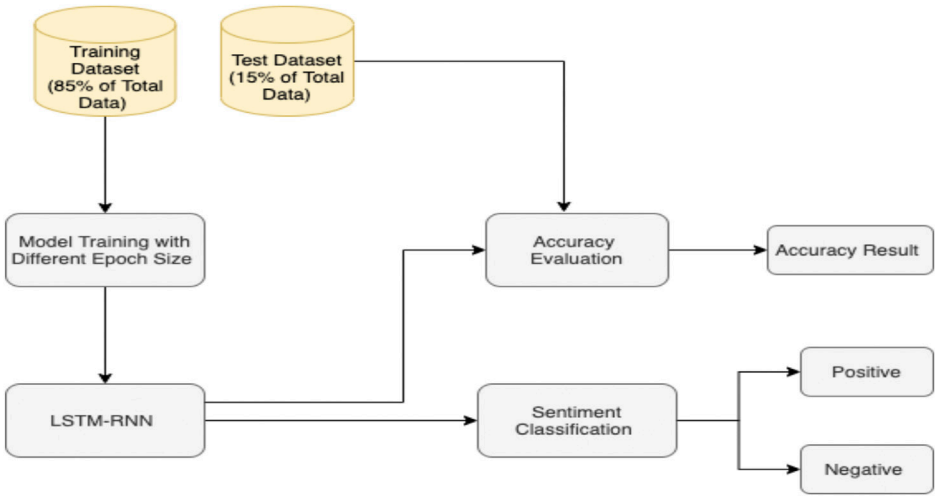


FIGURE 8 | Architecture of LSTM result.

TABLE 1 | Comparison of LSTM RNN Model Based on Epoch size.

Sectors	Epoch size	Loss	Training set accuracy (%)	Test set accuracy (%)	Validation accuracy (%)
Tourism Hospitality	5	0.225	91.1	80.3	79.7
	10	0.043	98.4	80.4	80.4
	15	0.011	99.6	80.9	80.9
Tourism Healthcare	5	0.168	93.7	78.1	78.1
	10	0.029	99.1	78.7	78.7
	15	0.007	99.3	76.1	77.7

country's economy tanking and the prime minister being held accountable.

A country like India, which is developing and has a lot of promise, may also be the reason why individuals began to tweet that shutting the country's borders at an early stage would have saved a lot of lives and prevented further spread of the virus.

Violin Plot of Compound Score for Each Sector

The violin plot of all the sectors is shown in **Figure 7**. A typical observation is that the violin plot of all positive sentiments has a higher degree of fluctuation, particularly in the hospitality sector. This means that the compound score from the range of 0.2–0.5 is higher, indicating that there are more tweets in that area. In these plots, the straight line denotes that the class belongs to neutral sentiment, which has a compound score of 0. If the violin plot has a wider region, as it does in the instance of negative emotions in the tourism economic sector, it indicates that the composite score of all negative sentiments follows a consistent range.

LSTM RNN Model Results

The input/output of the model with 85% of total data for training and 15% have been kept for testing purpose shown in **Figure 8**. **Table 1** summarizes the full experimental findings and comparisons of the LSTM RNN deep learning model for three epoch sizes: 5, 10, 15, and 20 for each tourism sector. To begin, an epoch size of 15 obtained the maximum accuracy of 80.9 percent and a loss of 0.011 in the tourist hospitality sector. Secondly, the model performed well in the tourism economy sector, with an accuracy of 75% and a loss of 0.043, which is 5% less than the preceding sector's performance. Furthermore, in the fifth and 10th epochs, the final sector which is healthcare obtained an accuracy of 76.1 percent. Lastly, with the increase in the epoch size of 20, we saw that the model started overfitting the dataset. Hence, we went for the optimal epoch size of 15 which indicates the robustness of the model.

Overall, the deep learning model performed admirably and accurately categorized most of the sentiments, suggesting that we could depend on the model for future sentiment classification predictions.

Table 2 compares the random forest and support vector classifier, two common machine learning algorithms. When we examine the findings of **Table 1** and **Table 2**, especially the test set accuracy, it is apparent that the deep learning algorithm performs well in the hospitality industry. The SVC algorithm, on the other hand, does have a little higher accuracy in the healthcare sector. Another thing to remember is that the test set accuracy refers to the tweets on which the model has not been trained. The LSTM RNN

TABLE 2 | Comparison of machine learning models.

Random forest classifier results					
Sectors	**Sentiment**	**Precision (%)**	**Recall (%)**	**F1-score**	**Accuracy (%)**
Tourism Hospitality	Negative	83.3	50	62	79
	Positive	77	94	85	
Tourism Healthcare	Negative	85	70	59	74.6
	Positive	72	75	82	
SVC Results					
Sectors	**Sentiment**	**Precision (%)**	**Recall (%)**	**F1-Score**	**Accuracy (%)**
Tourism Hospitality	Negative	83	71	77	83.9
	Positive	84	92	88	
Tourism Healthcare	Negative	80	66	72	79.6
	Positive	80	89	84	

TABLE 3 | Similar terms and inter cluster dependency.

Cluster number	**Healthcare top tweet terms**	**Hospitality top tweet terms**
C1	tourism	Covid
C1	hospitals	hospitality
C1	Hospital	Hotel
C1	Healthcare	tourism hotels
C1	People	Hotels
C1	Covid	industry
Terms similarity = “covid”		
C2	Healthcare	tourism
C2	Tourism	hospitality
C2	Economist	hotel
C2	Medical	videos
C2	Hospital	industry
C2	Hospitals	marketing
Terms similarity = “tourism”		
C3	Medical	hotel
C3	Tourism	covid
C3	Hospitals	hotels
C3	Healthcare	tourism
C3	Hospital	travel
C3	Best	destination
Terms similarity - tourism		
C4	Healthcare	tourism
C4	Tourism	video
C4	Education	travel
C4	Medical	hospitality
C4	manufacturing	hotels
C4	Food	storytelling
Terms similarity “tourism”		
C5	Bbcnews	video
C5	Coronavirus	tourism
C5	Hospital	hospitality
C5	Tourism	benefits
C5	Medical	industry
C5	Healthcare	corporate
Terms similarity “tourism”		

algorithm was chosen primarily because neural networks are a very powerful technique for text and image classification. One of the model's shortcomings is that it has no memory, which is an issue for sequential data such as text or time series. RNN solves this problem by introducing a feedback glance that acts as a memory. As a result, the model's previous inputs leave a trace. LSTM expands on this concept by including a short-term and long-term memory component. As a result, LSTM is an excellent tool for anything involving a sequence. Because the meaning of a word is determined by the words that came before it. Because of this, we chose the LSTM RNN algorithm and took advantage of its memory feature. Both approaches have a test accuracy that goes hand in hand, as we observed before when comparing the results acquired by machine learning and deep learning (**Tables 1**, **2**). This is primarily due to the size of our dataset. Because we were constrained by the Twitter API's request limits, LSTM RNN, a deep learning algorithm, requires a large amount of data to reach its full potential.

Cluster Inter-dependency Between Healthcare and Hospitality

In topic modeling, the study retrieved the information of common terms which are widely used during the Covid time. Firstly, the clusters of top tweets in healthcare and hospitality were extracted with the help of the LDA approach. It can be observed in **Table 3** that top similar terms used during pandemic times are “covid” and “tourism” and even other words such as hospitals and hospitality have been used maximum times by social media users. It can be seen in **Figures 3**, **4** that the clusters were five in numbers namely - Cluster 1, 2, 3, 4, and 5. The top six tweet terms were taken from both the keywords-based topic namely - healthcare and hospitality. The Inter Topic Distance Map was done with the help of multidimensional scaling. The table provides a panoramic view of similar terms in both sectors. This provides the food for thought of cluster interdependency calculation. Six out of the top 30 salient terms were selected in the table. The priority was given based on the maximum percentage of the salient terms.

Our approach for this research have been proposed with three aspect, first topic modeling which shows that what kind phrases are being used during the pandemic, and this approach will helps to find the trending issues related to tourism healthcare and hospitality. Second, we used a deep learning technique based on the LSTM RNN algorithm to classify twitter users' emotions. In the hospitality and healthcare sectors, one may utilize our model, which was trained on

real-world data, to assess public opinion and identify any problems that may emerge in any hot zones of a country where the government can assist, be it with medical supplies or tending to stuck tourists. Lastly, we tried to find the similar terms interdependency in the cluster which acts as an extra advantage to analyse what kind of words are maximally being used between the two sectors.

DISCUSSION

This research work aims to extract and analyze tweets about coivd-19 from tourism sub-domains healthcare and hospitality from all around the world. The Covid-19 has had the greatest influence on the tourism domain, and visitors who had planned a vacation to another nation away from their regular surroundings have been trapped inside their homes.

The authors began by performing sentiment analysis with a very distinctive approach called VADER, which has the largest benefit of requiring very little data cleansing since it recognizes the magnitude of the words. After the analysis, it is found that the positive sentiments were the highest in all of the sectors which briefly means that In the next few years, it will have a substantial and beneficial influence on worldwide tourism. It also offers the possibility of transforming our affected planet into a greener one. Furthermore, we used topic modeling to discover patterns such as word frequency and distance between words. A topic model clusters comparable feedback as well as phrases and expressions that appear frequently. With this knowledge, it is possible to immediately infer what each group of texts is discussing. Each document in the LDA model is seen as a conglomeration of themes from the corpus. According to the model, each word in the text is related to one of the document's subjects. Lastly, implementation of the state-of-the-art deep learning algorithm and testing the model with different epoch sizes helped in providing a robust sentiment prediction model.

Hospitality

The hospitality industry has taken huge hits in terms of revenue, and massive layoffs occurred because of Covid-19. By implementing topic modelling, it is seen that the restaurants, hotels, and resorts are among the highest sectors facing difficulty. The proposed deep learning model has achieved the highest values of test set accuracy of 80.9 percent, which is the maximum among the epoch sizes between 5 and 10. Apart from test accuracy, the authors captured the loss and training set accuracy of 0.011 and 99.6 percent, respectively. The hospitality sector has contributed to providing shelters for tourists with the alliance of government agencies during lockdown phase. Many hotels came forwards to facilitate the best services for covid victims during the quarantine phase. This has resulted in sharing of positive narrative by tourists when they are isolated during lockdown periods.

Healthcare

Healthcare to tourists is a critical component in which the best medical treatment is provided to foreign tourists and visitors. The tweets for tourist healthcare were collected from visitors who were trapped in various nations and how they handled and shared their experiences on social media data. On an epoch size of 10, the created deep learning model for the tourism sector can produce the best results. The study provided various implications for practice. The Twitter analysis can be done to check the impact of various other crises on tourism domains. The sentiment analysis using VADER can be a boon for many researchers. The tourism economy has witnessed a huge halt due to travel restrictions. The study has provided a panoramic view of the analysis during the coronavirus crisis which can help in paving the foundation for the implementation of efficient remedial measures to sustain after the covid-19 pandemic. The social media analysis helps the healthcare admirative agencies to analyze the tourist sentiments and prioritize the urgency patients due to covid complications.

CONCLUSION AND FUTURE SCOPE

Social media platforms are a medium for sharing opinions and thoughts. Analysis of these thoughts may help to get ready for upcoming travelers once the Covid-19 situations will normalize. The study has a future scope of research. A post-crisis study can also be done after the Covid-19 pandemic is over. This can help provide a birds-eye view of the overall impact of the coronavirus on tourism domains.

In view of the Covid-19 pandemic, Twitter data of tourism two sub-domains: hospitality and healthcare were selected and extracted accordingly. Analysis is done through sentiment visualization and deep learning algorithms. Sentiment visualization is presented with the help of bar and violin plots. This study provides a strategic practical framework to sustain during the Covid- 19 pandemic which revolves around the deep learning analysis of social media. Finally, the LSTM RNN model prediction has a number of applications, one of which is allowing government officials to monitor social media sites such as Twitter in order to better understand their citizens' sentiments so that they can make choices and act in the best interests of the country and people.

One of the paper's future scopes is to see whether the topic modeling suggestion can be expanded to other sectors like airline and economy to see what kinds of comparable words are impacting and influencing the healthcare and hospitality industries. We were limited in the quantity of data we could collect because of the Twitter API's request restriction, which was one of the study's constraints. The LSTM RNN approach would be considerably more robust with a much larger dataset.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

Conceptualization, RKM; methodology, RKM; validation, RKM, SU and JJ; formal analysis, RKM; investigation, RKM, SU, JJ, AN and NN; writing original draft preparation, RKM, SU, JJ, AN and NN; supervision, SU and JJ.

REFERENCES

Aburumman, A. A. (2020). COVID-19 Impact and Survival Strategy in Business Tourism Market: the Example of the UAE MICE Industry. *Humanit Soc. Sci. Commun.* 7 (1), 1–11. doi:10.1057/s41599-020-00630-8

Agrawal, S., Sarkar, S., Srivastava, G., Reddy Maddikunta, P. K., and Gadekallu, T. R. (2021). Genetically Optimized Prediction of Remaining Useful Life. *Sust. Comput. Inform. Syst.* 31, 100565. doi:10.1016/J.SUSCOM.2021.100565

Alaei, A. R., Becken, S., and Stantic, B. (2017). Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Trav. Res.* 58 (2), 175–191. doi:10.1177/0047287517747753

Alamanda, D. T., Ramdhani, A., Kania, I., Susilawati, W., and Hadi, E. S. (2019). Sentiment Analysis Using Text Mining of Indonesia Tourism Reviews via Social Media. *Int. J. Humanit. Arts Soc. Sci.* 5 (2), 72–82. doi:10.20469/ijhss.5.10004-2

Ayoub, A., Mahboob, K., Rehman Javed, A., Rizwan, M., Reddy Gadekallu, T., Haider Abidi, M., et al. (2021). Classification and Categorization of COVID-19 Outbreak in Pakistan. *Comput. Mater. Continua* 69 (1), 1253–1269. doi:10.32604/CMC.2021.015655

Bhattacharya, S., Reddy Maddikunta, P. K., Pham, Q.-V., Gadekallu, T. R., Krishnan S, S. R., Chowdhary, C. L., et al. (2021). Deep Learning and Medical Image Processing for Coronavirus (COVID-19) Pandemic: A Survey. *Sust. Cities Soc.* 65, 102589. doi:10.1016/J.SCS.2020.102589

Bonta, V., Kumaresh, N., and Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Ajcst* 8 (S2), 1–6. doi:10.51983/AJCST-2019.8.S2.2037

Foo, L.-P., Chin, M.-Y., Tan, K.-L., and Phuah, K.-T. (2020). The Impact of COVID-19 on Tourism Industry in Malaysia. *Curr. Issues Tourism*, 1–5. doi:10.1080/13683500.2020.177795110.1080/13683500.2020.1777951

Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., Ra, I.-H., et al. (20202020). Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. *Electronics* 9 (2), 274. doi:10.3390/ELECTRONICS9020274

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., and Khan, W. Z. (2021). An Ensemble Machine Learning Approach through Effective Feature Extraction to Classify Fake News. *Future Generation Comp. Syst.* 117, 47–58. doi:10.1016/J.FUTURE.2020.11.022

Hutto, C., and Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, Washington, USA, April 30, 2008, 8, 1, 216–225. Available at: https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

IATA, 2021. IATA COVID-19 Survey. (n.d.). Retrieved September 10, 2021, from, Available at: https://www.iata.org/en/publications/store/covid-passenger-survey/.

Impact of the Pandemic on Tourism – IMF F and D. (n.d.), 2021. Retrieved September 10, 2021, Available at: https://www.imf.org/external/pubs/ft/fandd/2020/12/impact-of-the-pandemic-on-tourism-behsudi.htm.

Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., et al. (2020). COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. *Front. Public Health* 8, 357. doi:10.3389/FPUBH.2020.00357

Iwendi, C., Mahboob, K., Khalid, Z., Javed, A. R., Rizwan, M., and Ghosh, U. (2021). Classification of COVID-19 Individuals Using Adaptive Neuro-Fuzzy Inference System. *Multimedia Syst.*, 1–15. doi:10.1007/S00530-021-00774-W

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., et al. (2017). Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey. *Multimedia Tools Appl.* 78 (11), 15169–15211. Available at: https://arxiv.org/abs/1711.04305v2. doi:10.1007/s11042-018-6894-4

Jeong, B., Yoon, J., and Lee, J.-M. (2019). Social media Mining for Product Planning: A Product Opportunity Mining Approach Based on Topic Modeling and Sentiment Analysis. *Int. J. Inf. Manage.* 48, 280–290. doi:10.1016/J.IJINFOMGT.2017.09.009

Kaushal, V., and Srivastava, S. (2021). Hospitality and Tourism Industry amid COVID-19 Pandemic: Perspectives on Challenges and Learnings from India. *Int. J. Hospitality Manage.* 92, 102707. doi:10.1016/J.IJHM.2020.102707

Khan, H., Srivastav, A., and Mishra, A. K. (2020). Use of Classification Algorithms in Health Care. *Big Data Analytics and Intelligence: A Perspective for Health Care*, Bingley, Emerald Publishing Limited, 31–54. doi:10.1108/978-1-83909-099-820201007

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An Introduction to Latent Semantic Analysis. *Discourse Process.* 25 (2–3), 259–284. doi:10.1080/01638539809545028

Manzoor, F., Wei, L., Asif, M., Haq, M. Z. u., and Rehman, H. u. (2019). The Contribution of Sustainable Tourism to Economic Growth and Employment in Pakistan. *Ijerph* 16 (19), 3785. doi:10.3390/ijerph16193785

Pandey, H., Mishra, A. K., and Kumar, D. N. (2019). Various Aspects of Sentiment Analysis: A Review. Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019, KNIT Sultanpur, June 2020, SSRN J, doi:10.2139/SSRN.3349011

Park, J. H., Lee, C., Yoo, C., and Nam, Y. (2016). An Analysis of the Utilization of Facebook by Local Korean Governments for Tourism Development and the Network of Smart Tourism Ecosystem. *Int. J. Inf. Manage.* 36 (6), 1320–1327. doi:10.1016/J.IJINFOMGT.2016.05.027

Poole, J., and Misrahi, T. (2020). *Travel & Tourism.Global Economic Impact Trends 2020*. London, UK, World Travel & Tourism Council. *June*, 1–20, Available at: https://wttc.org/Portals/0/Documents/Reports/2020/Global Economic Impact Trends 2020.pdf?ver=2021-02-25-183118-360.

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena* 404, 132306. doi:10.1016/j.physd.2019.132306

Srivastav, A., Khan, H., and Mishra, A. K. (2020, Advances in Computational Linguistics and Text Processing Frameworks. *Ad, in Comp. Elec. Engineer.* 217–244. doi:10.4018/978-1-7998-2772-6.CH012

Tatum, M. (2020). Will Medical Tourism Survive Covid-19? *BMJ* 370, m2677. doi:10.1136/BMJ.M2677

Thompson, K. (1968). Programming Techniques: Regular Expression Search Algorithm. *Commun. ACM* 11 (6), 419–422. doi:10.1145/363347.363387

Uğur, N. G., and Akbıyık, A. (2020). Impacts of COVID-19 on Global Tourism Industry: A Cross-Regional Comparison. *Tourism Manage. Perspect.* 36, 100744. doi:10.1016/J.TMP.2020.100744

UNCTAD (2021a). COVID-19 and Tourism - an Update. United Nations Conference on Trade and Development, Geneva, Switzerland, June 30 2021.

UNCTAD 2021b. Global Economy Could Lose over $4 Trillion Due to COVID-19 Impact on Tourism, Available at: https://unctad.org/news/global-economy-could-lose-over-4-trillion-due-covid-19-impact-tourism. Accessed on September 10, 2021.

UNCTAD. 2021c, Global Economy Could Lose over $4 Trillion Due to COVID-19 Impact on Tourism. Available at: https://unctad.org/news/global-economy-could-lose-over-4-trillion-due-covid-19-impact-tourism. Accessed on October 4, 2021.

UNWTO (2020). UNWTO World Tourism Barometer and Statistical Annex, August/September 2020. *UNWTO World Tourism Barometer* 18 (5), 1–36. doi:10.18111/WTOBAROMETERENG.2020.18.1.5

UNWTO World Tourism Barometer and Statistical Annex (2020). Eighth Asia & Oceania Congress of Endocrinology. November 30-December 4, 1986, Bangkok, Thailand. Abstracts 2020. *J. Med. Assoc. Thai* 69 (7), 1–100. doi:10.18111/WTOBAROMETERENG.2020.18.1.7

Webster, J. J., and Kit, C. (1992), Tokenization as the Initial Phase in NLP. Proceedings of the 14th conference on Computational linguistics, Nantes, France, August 23-August 28, 1992, 1106. doi:10.3115/992424.992434

Williams, D., Hornung, H., Nadimpalli, A., and Peery, A. (2021). Deep Learning and its Application for Healthcare Delivery in Low and

Middle Income Countries. *Front. Artif. Intell.* 4, 30. doi:10.3389/FRAI.2021.553987

Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., et al. (2020). Journal Pre-proof Prevalence of Comorbidities in the Novel Wuhan Coronavirus (COVID-19) Infection: a Systematic Review and Meta-Analysis. *Int. J. Infect. Dis.* 94, 91–95. doi:10.1016/j.ijid.2020.03.017

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* 31, 1235–1270. doi:10.1162/neco_a_01199

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mishra, Urolagin, Jothi, Neogi and Nawaz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/fpubh.2021.798034

Forecasting Dengue Hotspots Associated With Variation in Meteorological Parameters Using Regression and Time Series Models

*Seema Patil and Sharnil Pandya**

Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India

For forecasting the spread of dengue, monitoring climate change and its effects specific to the disease is necessary. Dengue is one of the most rapidly spreading vector-borne infectious diseases. This paper proposes a forecasting model for predicting dengue incidences considering climatic variability across nine cities of Maharashtra state of India over 10 years. The work involves the collection of five climatic factors such as mean minimum temperature, mean maximum temperature, relative humidity, rainfall, and mean wind speed for 10 years. Monthly incidences of dengue for the same locations are also collected. Different regression models such as random forest regression, decision trees regression, support vector regress, multiple linear regression, elastic net regression, and polynomial regression are used. Time-series forecasting models such as holt's forecasting, autoregressive, Moving average, ARIMA, SARIMA, and Facebook prophet are implemented and compared to forecast the dengue outbreak accurately. The research shows that humidity and mean maximum temperature are the major climate factors and exhibit strong positive and negative correlation, respectively, with dengue incidences for all locations of Maharashtra state. Mean minimum temperature and rainfall are moderately positively correlated with dengue incidences. Mean wind speed is a less significant factor and is weakly negatively correlated with dengue incidences. Root mean square error (RMSE), mean absolute error (MAE), and R square error (R^2) evaluation metrics are used to compare the performance of the prediction model. Random Forest Regression is the best-fit regression model for five out of nine cities, while Support Vector Regression is for two cities. Facebook Prophet Model is the best fit time series forecasting model for six out of nine cities. Based on the prediction, Mumbai, Thane, Nashik, and Pune are the high-risk regions, especially in August, September, and October. The findings exhibit an effective early warning system that would predict the outbreak of other infectious diseases. It will help the relevant authorities to take accurate preventive measures.

Keywords: dengue fever, climate change, machine learning, prediction, time series forecasting, regression model

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
K. Gokulnath,
VIT-AP University, India
Joseph Henry Arinze Anajemba,
Hohai University, China
Ebuka Ibeke,
Robert Gordon University,
United Kingdom

***Correspondence:**
Sharnil Pandya
sharnil.pandya@sitpune.edu.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 October 2021*
Accepted: *04 November 2021*
Published: *26 November 2021*

Citation:
Patil S and Pandya S (2021) Forecasting Dengue Hotspots Associated With Variation in Meteorological Parameters Using Regression and Time Series Models. Front. Public Health 9:798034. doi: 10.3389/fpubh.2021.798034

INTRODUCTION

Climate change is variations in climate variables such as temperature, humidity, precipitation, rainfall, wind speed, etc. Climate Change occurs due to natural activities such as variations in the sun, volcanic explosions, or human activities like the emission of carbon dioxide and other greenhouse gases that cause global warming. Infectious diseases are categorized into foodborne, airborne, waterborne, and vector-borne infectious diseases. Vector-borne infectious diseases are transmitted to humans by a microbe, called vectors, such as mosquitoes, ticks, flies, etc. Dengue is a vector-borne infectious disease carried by mosquito vectors that is most susceptible to meteorological conditions. According to WHO, this pandemic spreads over 128 countries across the globe and increased eight times over the last 20 years affecting 4.2 million people in the year 2020. **Figure 1** shows the effects of climate change on disease vectors.

For understanding the spread of dengue, studying climate change and its effects specific to the disease is necessary. Temperature, rainfall, humidity, wind speed are the significant meteorological factors for the spread of dengue fever. Identifying the relationship between variation in these climatic factors and dengue incidences helps to predict the disease outbreak more accurately. An association has been found between the climatic parameters and dengue incidences for the selected locations in the proposed system. Machine Learning plays a vital role in developing a predictive model to understand the influx of dengue. Previously different classification and regression techniques were implemented for the prediction of dengue outbreaks for different locations. Considering varied geographic topography with changing climatic conditions and frequent disease outbreaks in the past, there is a need for better and accurate predictive models for early surveillance systems and improved prevention strategies. The following points highlight the paper's significant contributions:

- To collect climate and dengue incidence data for the selected locations for the past 10 years
- To identify the correlation between variations in climatic parameters and dengue incidences
- To implement various predictive models and show a comparative analysis based on different evaluation metrics
- To predict different climatic regions at risk in the future based on its climatic conditions

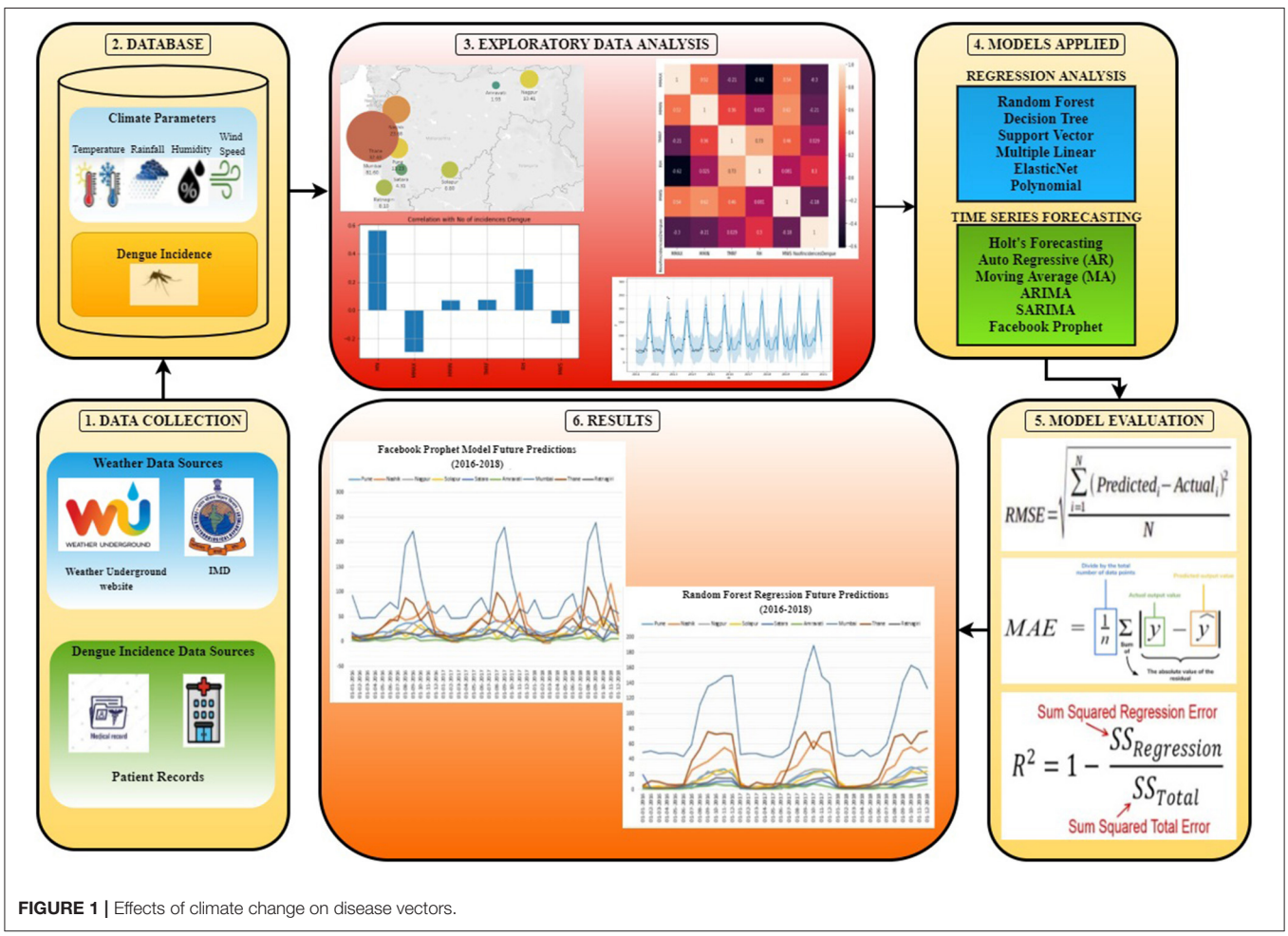


FIGURE 1 | Effects of climate change on disease vectors.

The following sections of the paper are organized as follows: Section Related Work describes related research work carried out for identifying the relationship between climate factors and vector-borne infectious diseases along with existing predictive modeling techniques available and its limitations. Section Proposed Work narrates the proposed system for dengue forecasting with variation in climate change. Section Methodology exhibits the methodologies used for forecasting dengue disease outbreaks. It includes different subsections such as data collection and integration, data preprocessing, exploratory data analysis, model execution, and evaluation metrics. Section Results and Discussion discusses details of predictive analysis and results. Finally, Section Conclusion, Limitations, and Future Work presents and concludes the author's research work.

RELATED WORK

Significant research has been carried out for understanding the association of meteorological variables with dengue incidences. This section describes the existing work related to the prediction of dengue incidences based on climatic factors using several machine learning techniques with its advantages and limitations.

Salim et al. (1) proposed a study to predict dengue outbreaks based on weekly dengue incidence data for the Selangor state of Malaysia. Several Machine Learning algorithms such as CART, ANN, SVM, and Naive Bayes create a predictive model. It has been found that the support vector machine model (SVM) best predicted dengue outbreaks. This research provides categorical output instead of continuous output. Liu et al. (2) implemented a unique approach for forecasting dengue incidences in Guangzhou, China. They integrated SVM-MLP machine learning approaches (3) with environmental features such as water collection sites, dustbins, etc. It performs better than models based on standard features (Temperature) alone. More standard features in addition to temperature and rainfall could be considered for better training of the ML Model. The SVR-based model Tanawi et al. (4) is proposed to predict dengue incidences in DKI, Jakarta. They concluded that SVR with a linear kernel provides better results than SVR with a radial kernel. Recently Mudele et al. (5) proposed a technique that uses a recurrent neural network (RNN) for forecasting the dengue mosquito vector population. This model is compared with random forest and k nearest neighbor for two Brazilian cities. They proposed that other deep learning models should be considered for the study (6–15). Mohapatra et. al. (16) investigated the effect of climate parameters on malaria outbreak using multilayer Perceptron and J48 classifier using WEKA tool. The results show that J48 is the most suitable model than MLP and has better accuracy and less error (RMSE). Also, temperature and humidity are more significant climate parameters than rainfall, and monsoon and post-monsoon are the peak periods for the outbreak. However, other factors such as demography, immunity within the population, society's socio-economic structure, availability of affordable public health facilities are not considered during the research (17). Cheng et al. (18) proposed distributed lag non-linear model to investigate the association between extreme weather events such as floods, heatwaves, high humidity, and dengue epidemic. The researcher implemented the model on daily dengue incidences and climate

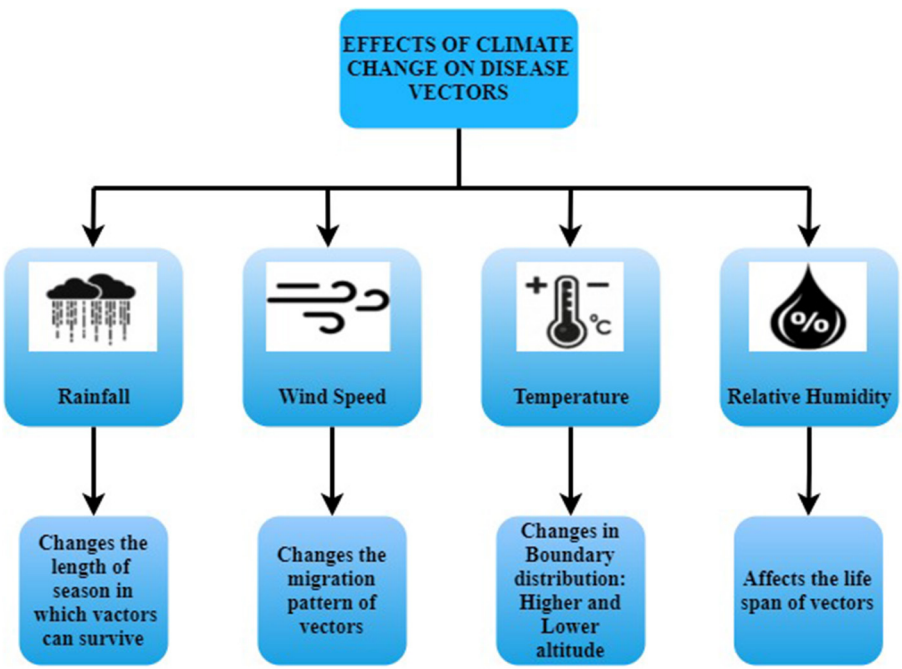


FIGURE 2 | Schematic overview of the proposed system.

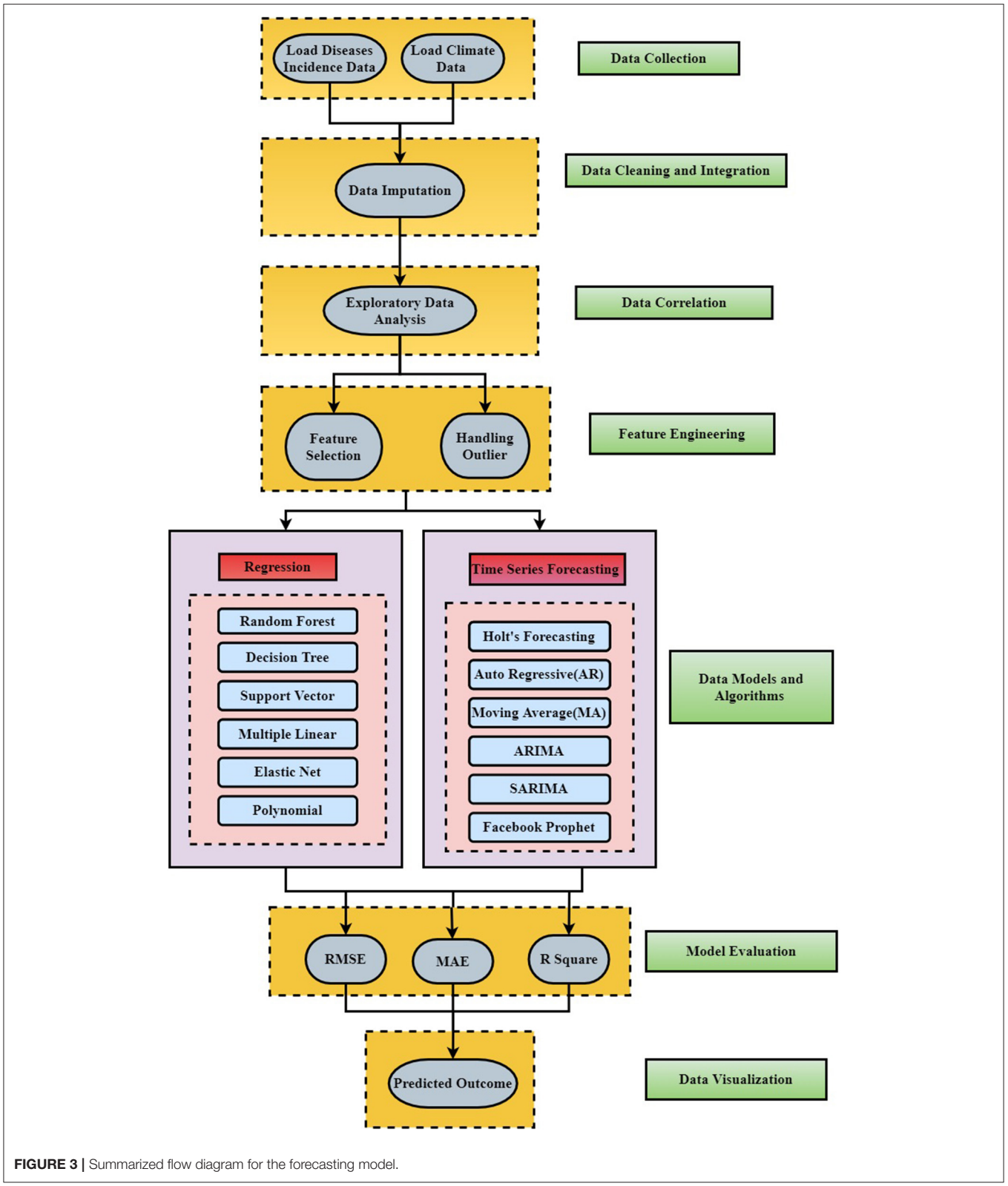


FIGURE 3 | Summarized flow diagram for the forecasting model.

factors such as temperature, humidity, and rainfall for different cities of China. The threshold for each climate parameter is calculated, and risk for dengue outbreak is identified for the extreme weather events. The limitation of the research is that

other time-variant factors such as changes in mosquito density, population movements and habits, and vector control measures are not considered for the study (19).

Xu et al. (20) analyzed dengue incidences data considering different meteorological factors. They proposed long short-term memory (LSTM) based recurrent neural network predictive model to predict monthly dengue cases using climate data for 20 Chinese cities. LSTM model shows the best performance for forecasting dengue incidences. But it is time-consuming compared to other models such as the backpropagation neural network and gradient boosting machine model. Appice et al. (22) formulated different strategies such as Auto Encoding, Window-based Data Slicing, and Cluster Analysis to discover temporal dynamics in temperature and dengue variables. They proposed a new multi-stage Machine Learning model called AutoTiC-NN (22) to find trend patterns between historical data of temperature and dengue in Mexico. The study proved that the model outperforms both in regression and time series forecasting analysis. Benedum et al. (23) compared machine learning, regression, and time-series models to forecast dengue cases and outbreaks in Peru, Puerto Rico, and Singapore. They concluded that Random Forest regression provided better results than Poisson Regression and ARIMA for short-term predictions while ARIMA was better for long-term forecasts. Nkiruka et al. (24) proposed a malaria incidence classification model (MIC) using climate parameters for six countries of Africa over 28 years. The research used k means clustering for outlier detection and the XGBoost model for classification. The proposed model is compared with other classification models such as ARIMA, SARIMA, SVM and showed the best results compared to other models.

Anno et al. (21) have integrated Spatiotemporal Hotspot analysis, RS Data, and a Machine Learning approach to develop a climate-based forecasting model to deliver early warning messages to the relevant public health authorities in Taiwan. This study uses two climate parameters (Rainfall and Temperature) to predict dengue outbreaks. Stolerman et al. (25) provide a better understanding of the long-term effects of climate conditions on the Aedes Aegypti (dengue causing mosquito) population. They have developed a new data-driven method using SVM algorithms to identify climate signatures that predict Dengue epidemics in Brazil. This research uses the binary threshold to classify epidemics/non-epidemics based on the Brazilian Ministry of Health. Two climate parameters (Frequency of precipitation and average Temperature) are used. Carvajal et al. highlighted the use of time lags of meteorological factors to predict dengue incidences. They concluded that Tree based Machine Learning methods (Random Forest, Gradient Boosting) performed better than conventional statistical techniques (GAM, SAIMAX) to predict a temporal pattern of Dengue incidences in Manila, Philippines. They also suggested that Relative Humidity is one of the most critical climate factors for their RF-LG model. All the variables are trained with keeping lag time in consideration to give an early outbreak prediction. Thus, this model cannot be used to predict an immediate output (17, 19, 26, 27). Despite continuous research, due to the varied topography of India, especially Maharashtra state having different climatic regions, there is a need to develop an accurate and enhanced predictive model for effective forecasting (2, 28–36). It will help the medical researcher and public health department promptly respond to the dengue outbreak and undertake corrective majors.

PROPOSED WORK

Figure 2 shows a schematic overview for dengue forecasting using regression and the time series model. It includes data sources and collection for both Climate parameters and dengue incidences. This is followed by data cleaning and integration in which missing data are imputed using the mean of the month data imputation technique. Exploratory data analysis is performed to find the correlation between climate parameters and dengue incidences. Feature engineering is carried out for feature selection and handling outliers. The impact of climate change includes indirect effects such as rising sea and temperature levels, extreme weather events such as droughts, floods, heatwaves, etc. The direct impact of climate change includes endurance, reproduction, or distribution of disease vectors which may affect human health. The climatic variations help in transmitting disease pathogens that may lead to infectious diseases (5, 16–19, 26, 27, 37–42).

Furthermore, the data is then split into training and testing data sets where training data is used to train different Machine Learning models–Regression Analysis and Time Series forecasting. These models are evaluated based on three evaluation metrics–Root Mean Square Error, Mean Absolute Error, and R Square Error. The models are compared to determine which models work best for different cities based on their geographic locations. Finally, locations at risk and outbreak period are predicted. Various visualization tools and techniques are used to represent the data and results effectively.

Novelties and Contribution of the Proposed Work

The effect of the variation in climate factors with varied topography on infectious diseases such as dengue is an exciting research area. The proposed work illustrates the detailed analysis of the climate and health data for different locations of Maharashtra state of India. It includes finding a correlation between monthly climate factors such as mean minimum temperature, mean maximum temperature, mean wind speed, relative humidity, etc., with dengue incidences for different locations. These locations have diverse geographic topography and weather conditions. Based on the analysis, forecasting of dengue outbreaks is performed using time series and regression models. The performance of these models is compared using various evaluation metrics and identifies the best suitable models for the study. This research will help design an effective surveillance system that will accurately monitor and control the dengue outbreak in a timely manner.

Methodology

Figure 3 shows the detailed workflow and layered architecture for the construction of the dengue forecasting model. The following sub-section (Data Sources and Collection, Data

Cleaning and Integration, Exploratory Data Analysis, Feature Engineering, Model Execution, Model Evaluation) elaborates different data sources and data collection process along with data preprocessing techniques implemented such as data imputation for missing values, climate and health data integration, feature selection, and outlier detection. Exploratory data analysis is performed to identify the correlation between climate parameters and dengue incidences using different visualization techniques such as heat maps, feature plots, etc. It also depicts different time series and regression machine learning models applied along with model evaluation metrics. Finally, dengue outbreak is forecasted for different cities of India for the next 3 years.

TABLE 1 | City wise population and weather conditions.

Division name	City	Population (in Lakhs)	Weather conditions
Konkan	Mumbai	125	Tropical, wet, and dry climate
Konkan	Thane	18.9	Tropical monsoon climate
Konkan	Ratnagiri	3.27	Tropical
Pune	Pune	31.2	Hot, semi-arid climate
Pune	Solapur	9.51	Dry, arid, and semi-arid climate
Pune	Satara	3.26	Tropical, pleasant climate
Nashik	Nashik	14.86	Mild tropical climate
Nagpur	Nagpur	24.1	Tropical Savanna climate
Amaravati	Amaravati	6.47	Tropical, wet, and dry climate
Aurangabad	Parbhani	3.07	Tropical, hot, and wet climate

Data Sources and Collection

Climate Data

Maharashtra has diverse climatic regions like Kokan, Khandesh, Desh, Vidarbha, and Marathwada. Based on the intensity of the disease and varied climatic conditions, nine cities like Mumbai, Thane, Ratnagiri, Pune, Solapur, Satara, Nashik, Nagpur, and Amaravati have been selected for the study. Monthly climate data is collected from Indian Meteorological Department (IMD) for 10 years from 2009 to 2019. The parameters in consideration are attributes such as Monthly Mean Maximum Temperature (MMAX) (°C), Monthly Mean Minimum Temperature (MIN) (°C), Total Monthly Rainfall (TMRF) (mm), Relative Humidity (RH) (%), and Mean Wind Speed (MWS) (km/h). **Table 1** shows different region-wise locations of Maharashtra state along with population and climatic conditions.

Health Data

The monthly dengue disease incidence data is collected from the National Vector Borne Disease Control Program (NVBDCP) for targeted cities of Maharashtra state mentioned in the climate data section for 10 years from 2009 to 2019. The data collected is in excel format, having inconsistent and missing values. The climate and disease incidence data are integrated into the CSV file for all the nine targeted cities, and data preprocessing is performed. **Figure 4** shows the map of Maharashtra state with region-wise selected cities for the study.

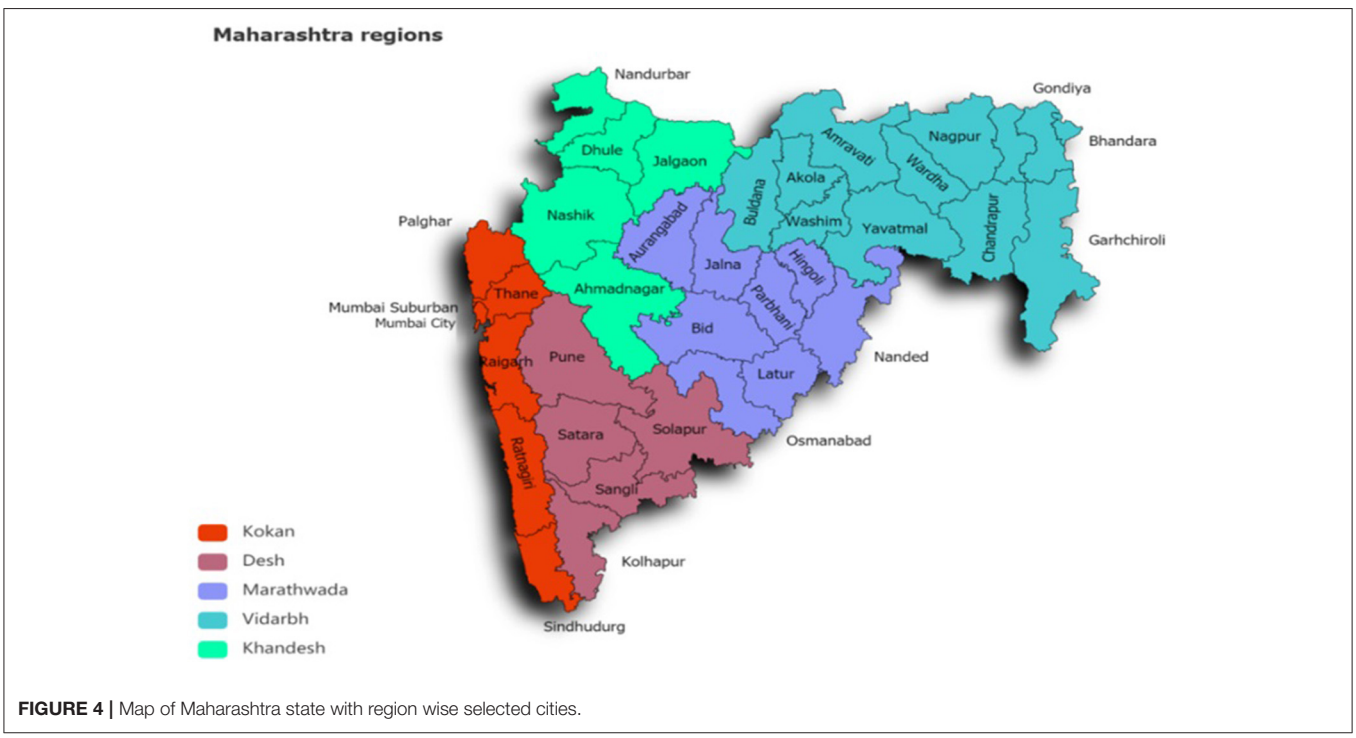


FIGURE 4 | Map of Maharashtra state with region wise selected cities.

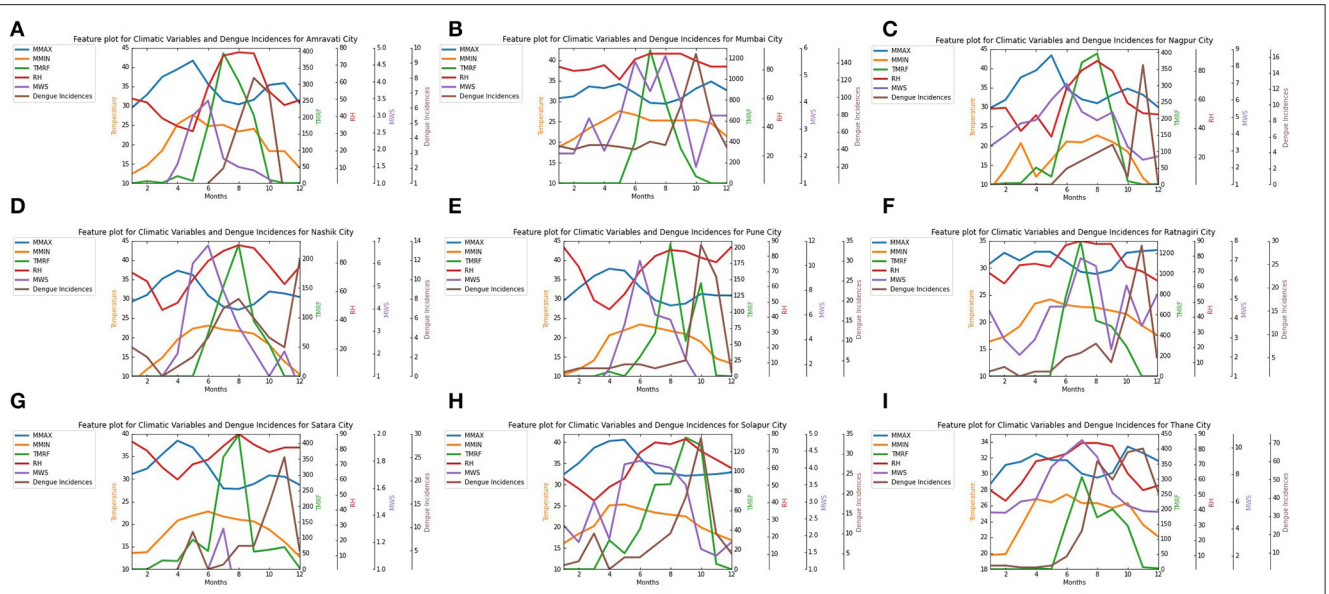


FIGURE 5 | City Wise Features Plot for climate variables monthly mean minimum temperature, mean maximum temperature, Average rainfall, Relative humidity, Mean wind speed, and monthly dengue incidences for nine selected cities in Maharashtra from 2009 to 2019. **(A)** Amravati, **(B)** Mumbai, **(C)** Nagpur, **(D)** Nashik, **(E)** Pune, **(F)** Ratnagiri, **(G)** Satara, **(H)** Solapur, **(I)** Thane.

Data Cleaning and Integration

To create the dataset, climate and dengue incidences data are collected and integrated. The dataset generated had inconsistent values due to the diverse nature of weather and health data. For each targeted city, a few irrelevant attributes are removed from the dataset during integration. The resulting dataset consists of missing values, especially in climate parameters. Data cleaning is performed to identify missing values. The data imputation technique is used to clean the dataset. The missing data were imputed using the mean of the Month Imputation technique. In this method, the missing values are replaced with an average of the previous values of the same month throughout different years. The mean of the month imputation function is given by:

$$\text{Vest} = (\sum \text{Tj} = 1\text{Vij})/\text{T} \quad (1)$$

$$\text{Vest} = (\text{Vij1} + \text{Vij2} + \text{Vij3} + \ldots.. + \text{VijT})/\text{T} \quad (2)$$

The estimated value Vest for the missing attribute is calculated by the averaging sum of values (Vij) of the variable for the ith month of the year j, where T is the number of available data for that year. In the present study, the mean of Maximum temperature "MMAX" for August 2016 was missing in the given dataset from 2009 to 2019. The estimated value is calculated by an average of previous values of the same month throughout different years. This value was treated as a data point in place of the missing value.

Exploratory Data Analysis

Once the dataset is cleaned, exploratory data analysis is performed to analyze attributes and summarize its characteristics using statistical techniques to discover useful patterns and graphical representation. City-wise feature graphs are plotted as shown in **Figures 5A–I**, and it is determined that each parameter for every city has a lot of variations, and there is no fixed pattern. So the correlation between each climate parameter and dengue incidences is found for all targeted locations to check which parameters are more significant.

Pearson correlation is performed on the dataset to determine the association between climate variables and dengue incidences, and heat maps are generated for each targeted city. Pearson correlation is a parametric test that measures the degree of relationship between two variables. It is the most suitable correlation technique based on the method of covariance and deals with numeric values. The person correlation function is given by Manogaran and Lopez (7):

$$r = \frac{\sum_{i=1}^{n} (ai - a)(bi - b)}{\sqrt{\left[\sum_{i=1}^{n} (ai - \overline{a})^2\right]\left[\sum_{i=1}^{n} (bi - \overline{b})^2\right]}} \quad (3)$$

Here, the Pearson correlation coefficient function is employed to determine the relationship between the climate parameters and the number of dengue cases. Climate variables are monthly mean max temperature (MMAX), mean minimum temperature (MMIN), Rainfall (TMRF), Relative Humidity (RH), Mean Wind Speed (MWS).

Correlation between climate factors and dengue incidences shows that each climate variable affects the dengue incidences differently. The mean maximum temperature (MMAX) is negatively correlated with the incidences of dengue despite the locations. This implies that as the maximum temperature decreases, incidences of dengue have increased. Mean minimum temperature (MMIN) is weakly/moderately positively correlated

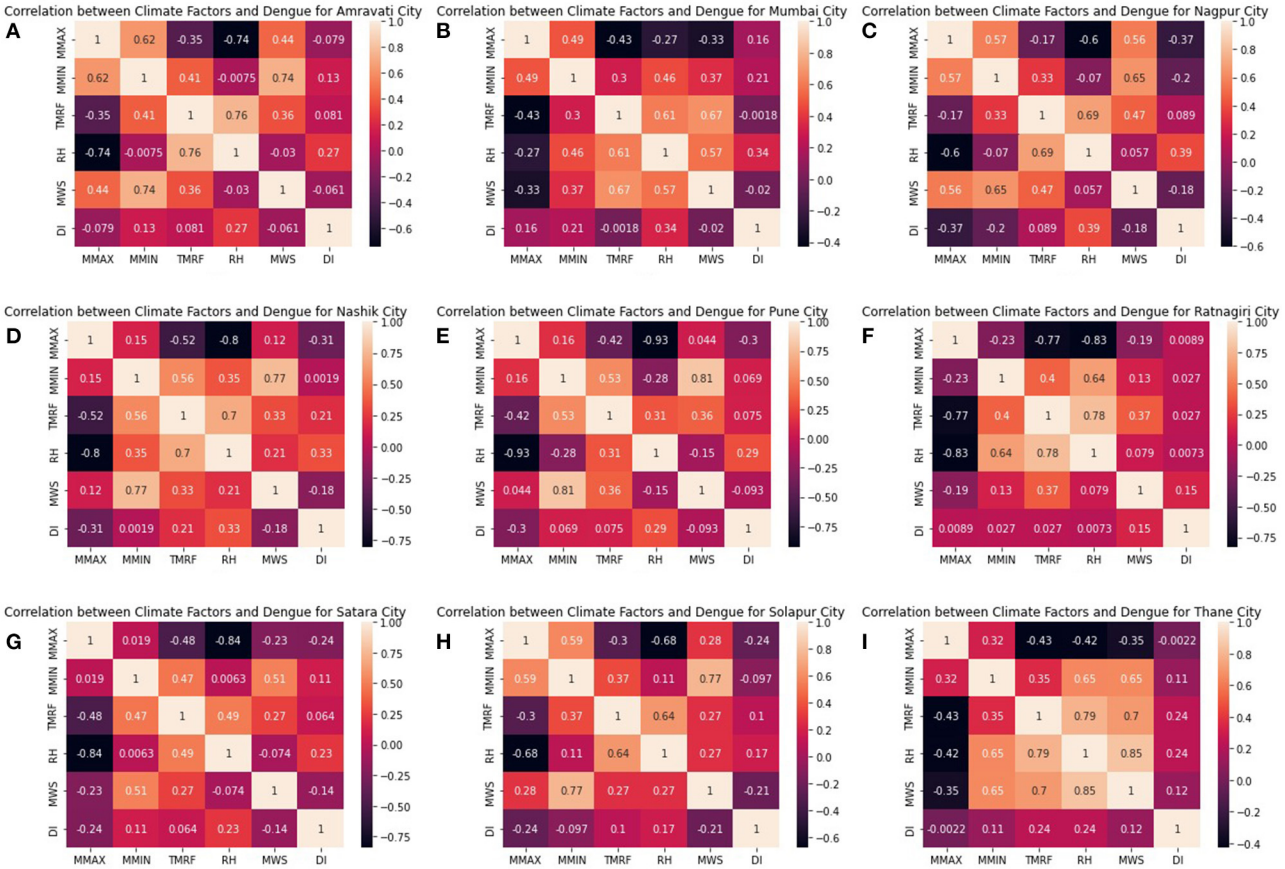


FIGURE 6 | Correlation of monthly mean minimum temperature, mean maximum temperature, total monthly rainfall, relative humidity, mean wind speed climate parameters, and monthly dengue incidences: **(A)** Amravati, **(B)** Mumbai, **(C)** Nagpur, **(D)** Nashik, **(E)** Pune, **(F)** Ratnagiri, **(G)** Satara, **(H)** Solapur, **(I)**Thane.

with dengue incidences except for Nagpur. Relative Humidity (RH) is the primary climate factor and exhibits a strong positive correlation with dengue incidences for all locations of Maharashtra state. Similarly, total monthly rainfall (TMRF) is moderately positively correlated with incidences of dengue. As humidity or rainfall is increased, cases have shown an increase for all selected cities of Maharashtra. Maximum incidences occur between June to September, where the average rainfall is between 150 and 350 mm. Mean Wind speed (MWS) is a less significant climate factor and weakly negatively correlated with dengue incidences.

Figure 6 shows city-wise graphs of the Pearson correlation of each climate parameter with the dengue incidences. These graphs are further used for feature selection based on results generated to identify the significant climates factors.

Feature Engineering

The data quality is of utmost importance for developing a predictive model with better accuracy and faster performance. For this purpose, a few data preprocessing techniques are applied, such as outlier detection and feature selection, to improve the data quality. The meteorological data consist of extreme values for specific periods, such as extreme wind speed, rainfall, and humidity. Outliers in the dataset can reduce predictive modeling performance. So the final dataset was normalized to uniform into the same scale.

Feature or attribute selection is the process of selecting the most relevant attributes in a dataset that helps train the model faster, reduces overfitting, and improves the accuracy of the predictive model. Minimal redundancy maximum relevance feature selection technique is used for attribute selection on the dataset to select attributes with high correlation and low variance. For determining the relevant features, two measures are calculated: redundancy and relevance. The following equation is used to find the mean of logical values of each climate parameter for the selected city in terms of dengue incidences:

$$\bar{c}_l = \frac{1}{n_i} \sum_{k=1}^{n_i} C_{ik} \tag{4}$$

Where, $\bar{c}_i$: Means of climate parameter i, n_i : Number of climate parameters, C_{ik}: the kth value of climate parameter i.

The below equation shows a variance of the climate parameters triggered by dengue incidences:

$$C_i^2 = \frac{1}{n_i} \sum_{k=1}^{n_i} (C_{ik} - \bar{c}_l)^2 \tag{5}$$

TABLE 2 | Dropped variables.

City	Dropped variables
Amravati	N/A
Mumbai	TMRF
Nagpur	N/A
Nashik	MMIN
Pune	TMRF, MMIN
Ratnagiri	MMAX, RH
Satara	TMRF
Solapur	N/A
Thane	MMAX

The minimal redundancy condition for the redundancy measure can be expressed by:

$$\min R(D),\ R = \frac{1}{|d|^2} \sum_{yi;yj \in d} I(yi; yj) \tag{6}$$

Where, min R (D) is the minimal redundancy for redundancy measure R, $|d|$ is the number of features in the subset of feature D, and I (yi; yj) is the mutual information between feature i and j.

The maximal relevance condition for the relevance measure can be expressed by:

$$\max RL(D, a),\ RL = \frac{1}{|D|} \sum_{yi \in d} I(yi; a) \tag{7}$$

Where max RL (D, a) is the maximal relevance for relevance measure RL and target activity a and I (yi; a) is the mutual information between the feature i and target activity a.

The smaller the value of the redundancy measure, the better the criteria for selection. Similarly, the higher the value of relevance measure, the better the feature selection. After exploratory data analysis is performed on the dataset considering several feature variables such as MMIN, MMAX, TMRF, RH, and MWS, different climate variables with high redundancy and low relevance are dropped for few cities under study, as shown in **Table 2**.

Model Execution

Regression is a supervised learning statistical method used to estimate the relationship between a dependent and one or more independent variables to determine trends in the data. It is used in the prediction of continuous values. The following regression models are implemented to predict dengue incidences across different cities based on climate parameters in the proposed system.

Support Vector Regression

It is the regression technique used to predict continuous ordered values. Some commonly used keywords in SVR are the kernel, hyperplane, boundary line, and support vectors. The primary purpose of SVR is to consider as many data points as possible within the boundary lines, and the hyperplane (best-fit line) must contain as many data points as possible. It is easy to implement and shows high prediction accuracy with excellent generalization capability. It can handle outliers very well.

Multiple Linear Regression

It is an extension of simple linear regression that models a linear relationship between more than one independent variable and a single dependent continuous variable. It is a technique for fitting a regression line through a multidimensional space of data points.

ElasticNet Regression

Elastic net is a type of regularized linear regression that includes two well-known penalties, the L1 and L2 penalty functions. The advantage of the elastic net model is that it permits a balance of both penalties, resulting in a more excellent performance on particular tasks than a model with either one or more penalties.

Polynomial Regression

Polynomial regression is a type of linear regression that estimates the connection as an nth degree polynomial. It is an example of multiple linear regression. Because Polynomial Regression is sensitive to outliers, the existence of one or two of them can have a negative impact on the results.

Decision Tree Regression

It is a regression model that breaks down a dataset into smaller subsets forming a tree with decision nodes and leaf nodes. Decision trees are very easy to visualize and reduce the uncertainty in the prediction. However, overfitting and underfitting are common problems with decision trees. If the hyperparameters are incorrectly set, the decision tree's output can vary dramatically.

Random Forest Regression

Random forest is the most commonly used machine learning technique that gives excellent results in predicting disease incidences based on climate conditions. It comprises many decision trees, each with the same node but different inputs, resulting in various leaves. It combines the results of the average of various decision trees. Overfitting can be avoided in the model by using Random Forest regression to create random subsets of the dataset.

Along with regression, the proposed system also used time series forecasting models to predict dengue incidences. Time series data is a sequence of different data points that measure a specific variable over an ordered period. In this method, time-series data extract meaningful statistics and other data characteristics to generate forecasts of our target variable. Different time series forecasting models are applied as given below:

Holt's Forecasting: It is time series forecasting method that depicts trends and seasonality from historical data. It is simple to implement and evolve with changing business requirements.

Auto-Regressive (AR) and Moving Average (MA): Autoregression forecasting technique predicts future values using previous values in time series. It demonstrates linear relation between future and past values. It is used to forecast recurring patterns in the data. The moving average method uses an average of several past points to predict future points. In this

method, short-term fluctuations and the effect of extreme values are reduced.

Auto-Regressive Integrated Moving Average (ARIMA): ARIMA model is a popular time series forecasting model which uses its lags to predict future values. It uses dependent relation between past observation and current observation. It involves subtracting recent observations from previous period observations several times. It is broken down into its subtypes to increase the accuracy of incidence predictions based on climate variability. This model does not support seasonal data.

Seasonal Auto-Regressive Integrated Moving Average (SARIMA): When seasonal components are added to the ARIMA model, then it is called SARIMA. It supports univariate time series data. Additional four seasonal elements in SARIMA are P, D, Q, and m, where P is seasonal autoregressive order, D is seasonal difference order, Q is moving average order, and m is the number of time steps (17).

TABLE 3 | Rudimentary observations.

Climate parameters	Rudimentary observations
Relative humidity	For RH > 55%, Number of incidences increases
Monthly maximum temperature	For temp 30°C < MMAX < 35°C, Number of incidences increases
Monthly minimum temperature	For temp 18°C < MMIN < 25°C, Number of incidences increases
Mean wind speed	For MWS < 2.5 kmph, No of incidences increases
Total monthly rainfall	For rainfall, 150 mm < TMRF < 250mm, No of incidences increases

Facebook Prophet Model

The Facebook prophet is a relatively new time series forecasting model developed in 2017 by the Facebook data science team as open-source software. In this model, irregular observations are permitted in the dataset as it ignores temporal data dependence. It is accurate, fast, and shows excellent performance as compared with other time series forecasting models. The prophet equation is given by:

$$x(t) = c(t) + s(t) + h(t) + u(t) \tag{8}$$

Where x(t) is forecast value, c(t) is the trend, i.e., change over a long period, s(t) is the seasonality, h(t) is the effect of the

TABLE 4 | Performance metrics comparison table for Amravati.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	5.22	3.43	−0.32	Holt's Forecasting	2.22	2.09	−0.39
Decision tree	5.35	2.88	−0.36	AR	1.95	1.41	−0.07
Support vector	4.82	**2.11**	−0.12	MA	**1.93**	**1.29**	−0.05
Multiple linear	4.52	3.12	0.01	ARIMA	2.19	1.39	−0.35
ElasticNet	**4.49**	3.05	**0.02**	SARIMA	1.98	1.92	−0.11
Polynomial	7.44	5.7	−1.68	Facebook Prophet	3.14	2.1	**0.38**

TABLE 5 | Performance metrics comparison table for Mumbai.

Regression models	RMSE	MAE	R^2	Time Series forecasting	RMSE	MAE	R^2
Random forest	47.93	34.93	−0.28	Holt's forecasting	127.2	110.9	−2.49
Decision tree	**28.44**	**19.5**	**0.54**	AR	148.5	145.9	−3.76
Support vector	47.55	26.7	−0.26	MA	54.46	50.04	0.35
Multiple linear	60.5	45.25	−1.05	ARIMA	94.62	93.65	−0.93
ElasticNet	55.53	43.11	−0.72	SARIMA	55.7	54.68	0.32
Polynomial	61.01	46.12	−1.08	Facebook prophet	**23.27**	**15.57**	**0.84**

TABLE 6 | Performance metrics comparison table for Nagpur.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	**13.56**	**9.72**	**0.12**	Holt's forecasting	7.71	5.95	−0.82
Decision tree	14.89	10.61	−0.05	AR	**5.5**	**4.28**	0.07
Support vector	17.59	10.94	−0.46	MA	7.49	6.09	−0.71
Multiple linear	15.23	11.02	−0.1	ARIMA	5.52	4.43	0.06
ElasticNet	16.17	11.51	−0.24	SARIMA	12.87	11.83	−4.07
Polynomial	29.54	20.57	−3.13	Facebook prophet	9.9	7.28	**0.56**

TABLE 7 | Performance metrics comparison table for Nashik.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	**24.16**	**18.88**	**0.21**	Holt's forecasting	19.99	17.69	−3.63
Decision tree	44.55	29.5	−1.68	AR	11.95	10.59	−0.66
Support vector	33.36	21.18	−0.5	MA	15.36	13.33	−1.73
Multiple linear	31.32	24.07	−0.32	ARIMA	**11.72**	10.28	−0.59
ElasticNet	30.2	22.94	−0.23	SARIMA	12.17	**7.86**	−0.72
Polynomial	73.39	50.83	−6.27	Facebook prophet	23.27	16.19	**0.524**

TABLE 8 | Performance metrics comparison table for Pune.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	**14.4**	**9.44**	**0.25**	Holt's forecasting	18.25	17.73	−1.3
Decision tree	25.99	18.11	−1.43	AR	21.94	19.63	−2.32
Support vector	20.59	13.28	−0.53	MA	34.07	32.42	−7.02
Multiple linear	14.93	11.09	0.19	ARIMA	25.99	24.24	−3.67
ElasticNet	14.78	10.99	0.21	SARIMA	24.23	21.6	−3.05
Polynomial	18.51	12.21	−0.23	Facebook prophet	**9.3**	**6.7**	**0.64**

TABLE 9 | Performance metrics comparison table for Ratnagiri.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	**9.35**	7.31	**0.06**	Holt's forecasting	6.09	5.8	−7.79
Decision tree	15.44	10.16	−1.54	AR	4.42	3.32	−3.62
Support vector	11.46	**7.11**	−0.4	MA	6.57	4.75	−9.24
Multiple linear	11.2	8.96	−0.33	ARIMA	4.49	3.91	−3.79
ElasticNet	10.8	8.47	−0.24	SARIMA	**3.22**	**2.95**	−1.46
Polynomial	18.14	13.12	−2.51	Facebook prophet	7.14	4.89	**0.49**

TABLE 10 | Performance metrics comparison table for Satara.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	**5.93**	3.9	**−0.04**	Holt's forecasting	19.99	18.07	−21.21
Decision tree	6.56	4.83	−0.27	AR	13.71	12.92	−9.44
Support vector	6.43	**3.78**	−0.23	MA	13.71	12.92	−9.44
Multiple linear	5.97	4.57	−0.06	ARIMA	13.71	12.92	−9.44
ElasticNet	5.98	4.61	−0.06	SARIMA	12.72	12	−8
Polynomial	16.47	13.11	−7.06	Facebook prophet	**3.98**	3.2	**0.64**

holiday, u(t) is unconditional changes or error. This model works best with time series with substantial seasonal influences and historical data from multiple seasons. It is robust to outliers and handles missing values very well. This model gives the best performance for six out of nine cities to forecast dengue incidences based on climate variations in the proposed system.

Model Evaluation

Once all regression and time series forecasting models are trained, the performance of the models is evaluated using three evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R Square Error (R^2). RMSE is the Standard Deviation of predicted errors. Lower RMSE values indicate better models. RMSE is evaluated by the Equation (2):

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{\mathrm{N}} (\mathrm{x_t} - \bar{\mathrm{x}}_\mathrm{t})^2 / \mathrm{N}} \tag{9}$$

Here, xt is actual dengue incidences for time t and $\bar{x}_t$ is the predicted number of incidences by the model.

Mean Absolute Error (MAE) is the difference between the actual values and the predicted values. Lower MAE values indicate better models. MAE is evaluated by Equation (1):

$$MAE = 1/n \sum_{i=1}^{n} |\mathrm{X_i} - \mathrm{X}| \tag{10}$$

TABLE 11 | Performance metrics comparison table for Solapur.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random forest	13.07	9.75	−1.82	Holt's forecasting	18.46	17.66	−12.7
Decision tree	13.38	8.83	−2	AR	21.09	20.51	−16.88
Support vector	**8.2**	**5.14**	**−0.12**	MA	22.85	22.04	−19.99
Multiple linear	15.95	11.77	−3.27	ARIMA	21.09	20.51	−16.88
ElasticNet	15.31	11.34	−2.93	SARIMA	16.94	16.08	−10.53
Polynomial	23.64	18.46	−8.38	Facebook prophet	**7.67**	**5.25**	**0.55**

TABLE 12 | Performance metrics comparison table for Thane.

Regression models	RMSE	MAE	R^2	Time series forecasting	RMSE	MAE	R^2
Random Forest	34.4	27.77	−0.46	Holt's forecasting	14.65	13.19	−0.61
Decision Tree	35.67	27.16	−0.52	AR	12.88	**10.88**	−0.25
Support Vector	**33.67**	**24.49**	**−0.36**	MA	**12.47**	11.95	−0.17
Multiple Linear	34.12	26.19	−0.39	ARIMA	12.88	**10.88**	−0.25
ElasticNet	37.15	29.27	−0.65	SARIMA	20.13	16.6	−2.05
Polynomial	48.02	37.09	−1.77	Facebook prophet	21.88	15.97	**0.56**

R Square Error (R^2) is also known as the coefficient of determination. It tells us how well a model fits on a dataset. It indicates how close the regression line is to the actual data. The R^2 value closest to 1 is considered to be the best value. The equation given below evaluates the value of R^2 (20):

$$R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{Regression}}}{\mathrm{SS}_{\mathrm{Total}}} \tag{11}$$

RESULTS AND DISCUSSION

Based on the exploratory data analysis, it was observed that each climate variable affects the dengue incidences differently. The average temp range in Maharashtra state is between 26 and 43°C. As shown in **Figures 6A–H**, histogram graphs generated after performing Pearson's correlation shows that mean maximum temperature (MMAX) is negatively correlated with the incidences of dengue despite the locations. This implies that as the maximum temperature decreases, incidences of dengue have increased. Mean minimum temperature (MMIN) is weakly/moderately positively correlated with dengue incidences except for Nagpur. Relative Humidity (RH) is the primary climate factor and exhibits a strong positive correlation with dengue incidences for all locations of Maharashtra state. Similarly, total monthly rainfall (TMRF) is moderately positively correlated with incidences of dengue. As humidity or rainfall is increased, cases have shown an increase for all selected cities of Maharashtra. Maximum incidences occur between June and September, where the average rainfall is between 150 and 350 mm. Mean Wind speed (MWS) is a less significant climate factor and weakly negatively correlated with dengue incidences. **Table 3** shows rudimentary observations for all five climatic parameters. The performance of all the regression and time series forecasting models for each city is evaluated and compared.

Tables 4–12 present city-wise performance comparison for all regression and time series forecasting models. The best fit values for each metric are highlighted in bold.

Table 5 shows that decision tree regression gives the least values for RMSE, MAE, and R^2 compared to other regression techniques. At the same time, the Facebook prophet gives the least values for RMSE, MAE, and R^2 compared to other time series models for Mumbai city of Maharashtra. Similarly, **Table 6** shows that the random forest model gives the best values for all performance metrics, whereas the AR model gives the least values for RMSE (5.5) and MAE (4.28) for Nagpur city. **Table 7** shows that Random forest demonstrates the best performance for metrics RMSE (24.16), MAE (18.88), and R^2 (0.21) for Nashik city. **Table 8** shows that random forest gives the best performance for metrics RMSE (14.4), MAE (9.44), R^2 (0.25), and Facebook prophet gives the best performance for metrics RMSE (9.3), MAE (6.7), R^2 (0.64) for Pune city. From all the performance **Tables 4–12** and result analysis, it has been observed that Random Forest Regression is the best-fit regression model working on five out of nine cities, i.e., Nagpur, Nashik, Pune, Ratnagiri, Satara, whereas Support Vector Regression shows the best performance on two cities, Thane and Solapur. Facebook Prophet Model is the best fit time series model that worked on six out of nine cities in time series forecasting. For the rest of the cities, various combinations of ARIMA models worked as the best fit.

Figure 7 shows predictions for nine targeted cities using Random forest regression, and **Figure 8** shows the predictions using the Facebook prophet time series model for 36 months from the Year 2021 to 2023. A hot spot map of Maharashtra state is created, as shown in **Figure 9**, using Tableau to compare the average number of monthly cases across our selected cities to visualize these results. The figures show that Mumbai is the most affected city, with monthly average dengue cases going up to more than 80, while Amravati is

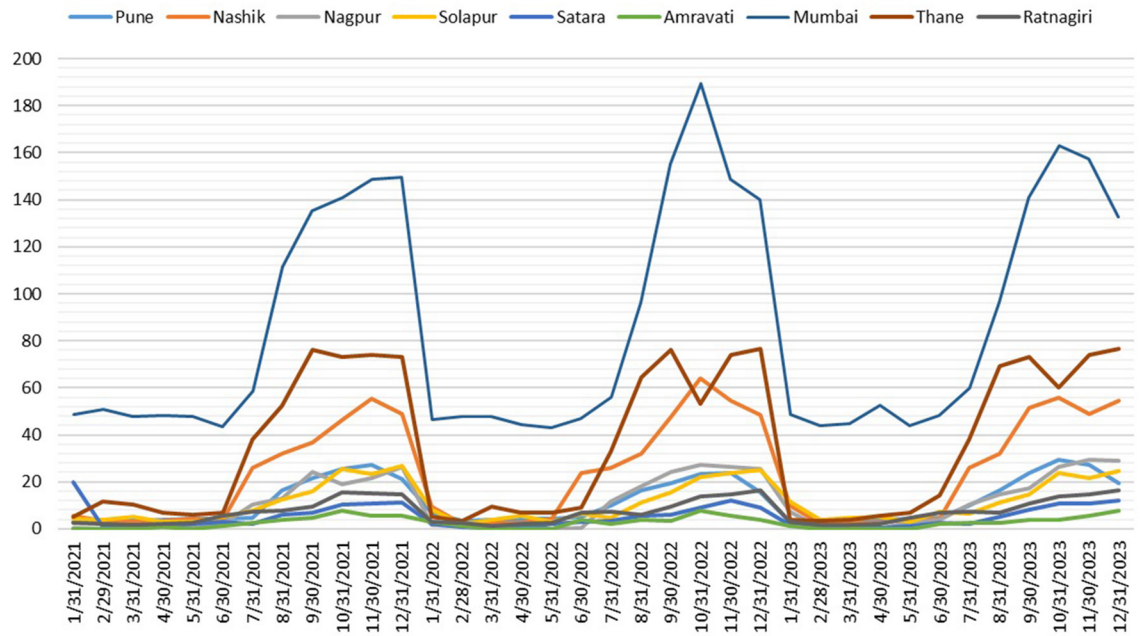


FIGURE 7 | Predictions for random forest regression model (X-axis: Date and Y-axis: Dengue cases).

the least affected location of Maharashtra. Other cities are ranged between 5 and 35 cases. Also, Thane, Nashik, and Pune are the cities at high risk, especially in August, September, and October.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

Conclusion

This research paper proposed a framework that can predict dengue incidences across different cities of Maharashtra based on climate parameters. Different meteorological variables like MMIN, MMAX, RH, TMRF, etc., are given as input, and the number of Dengue incidences is produced as output by the proposed system. Nine cities with varied climatic conditions were selected based on geographic regions. A correlation between meteorological parameters and dengue incidences was found out. The proposed system implemented 12 different regression and time series models for the prediction of dengue outbreaks. The performance of all the models is compared using root mean square error, mean absolute error, and R square error evaluation metrics. The result analysis shows that Random Forest outperforms the other Regression models for five out of nine cities. Facebook Prophet Model is the best fit time series forecasting model for six out of nine cities. The system also predicts the high-risk geographic regions from the year 2021 to 2023. It has been observed that Mumbai, Thane, and Pune are the hot spots in Maharashtra, especially from July to October. The medical researchers, public health departments, and health geography analysts can utilize these research results to take the necessary preventive measures based on these predictions.

Limitations

The study only considers climate factors. Non-climatic factors such as the demography, immunity within the population, society's socio-economic structure, availability of affordable public health facilities, and other environmental modifications initiatives are not considered for the study. Also, there is scope to add additional time-variant factors such as changes in mosquito density, population movements and habits, and vector control measures. The study is limited to a few cities of Maharashtra state of India to analyze monthly climate and dengue incidence data due to the unavailability of weekly or daily reports that could have helped better predictions.

Future Work

The result of the research will be helpful in designing an effective surveillance system that will effectively monitor and control dengue outbreaks. An output platform like a website can be created to assess the latest climate change parameters, disease outbreaks, and future projections. Future work can involve more extreme geographic regions of India along with daily or weekly

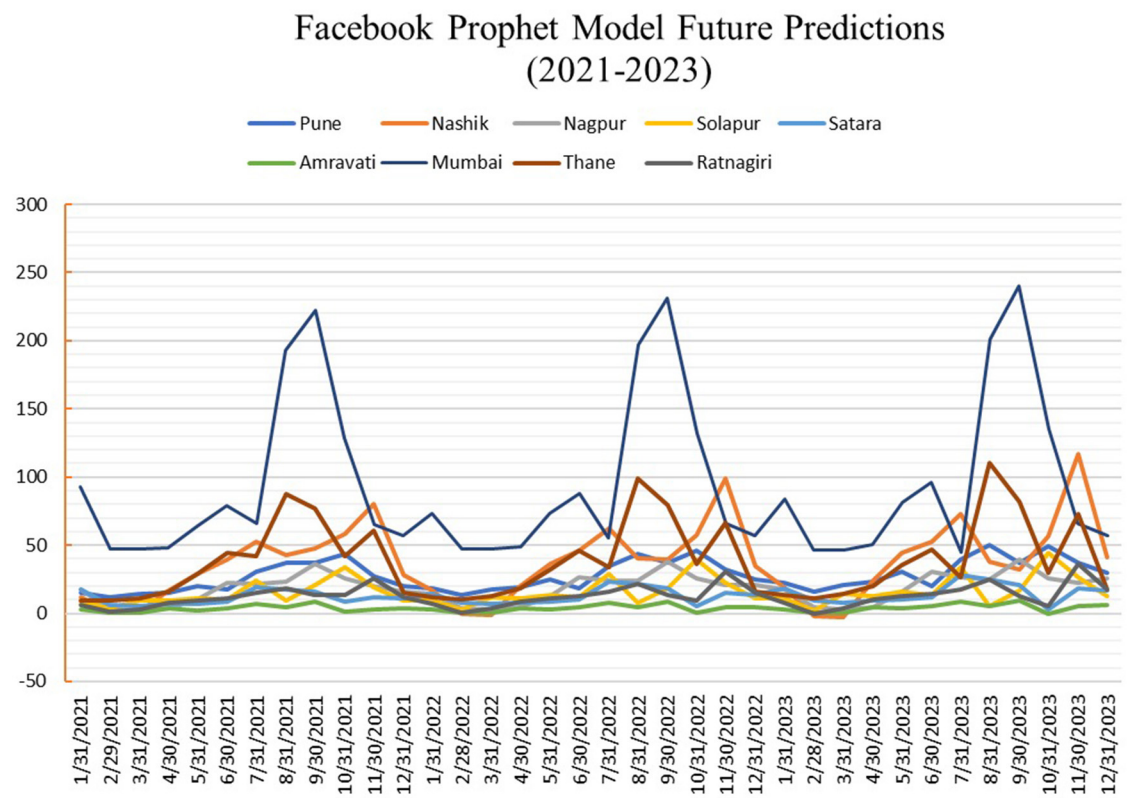


FIGURE 8 | Predictions for Facebook prophet model (X-axis: Date and Y-axis: Dengue cases).

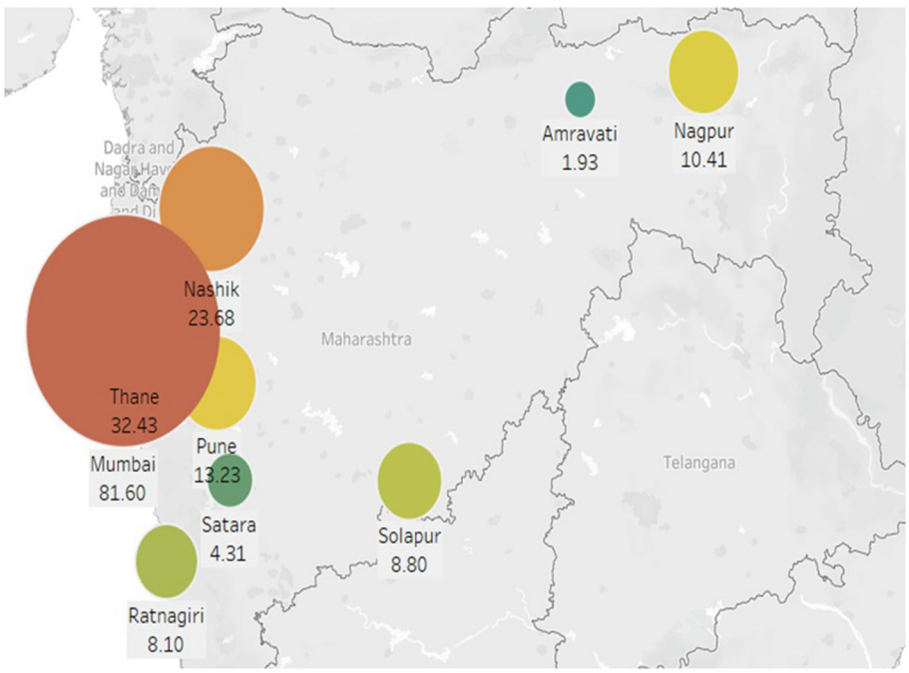


FIGURE 9 | Monthly average incidence hotspot map.

climate data analysis. Vulnerability groups such as age, gender, health status, occupation of the patients can be considered to enhance the surveillance system for better planning and preparation to avoid a future outbreak.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SPat and Span: conceptualization, data collection, interpretation, data curation, methodology, and manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

The research work has been supported by Symbiosis International (Deemed) University.

REFERENCES

1. Salim NAM, Wah YB, Reeves C, Smith M, Yaacob WFW, Mudin RN, et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Sci Rep*. (2021) 11:939. doi: 10.1038/s41598-020-79193-2
2. Semenza JC, Suk JE, Estevez V, Ebi KL, Lindgren E. Mapping climate change vulnerabilities to infectious diseases in Europe. *Environ Health Persp*. (2012) 120:385–92. doi: 10.1289/ehp.1103805
3. Liu K, Yin L, Zhang M, Kang M, Deng AP, Li QL, et al. Facilitating fine-grained intra-urban dengue forecasting by integrating urban environments measured from street-view images. *Infect Dis Poverty*. (2021) 10:40. doi: 10.1186/s40249-021-00824-5
4. Tanawi IN, Vito V, Sarwinda D, Tasman H, Hertono GF. Support vector regression for predicting the number of dengue incidents in DKI Jakarta. *Proc Comput Sci*. (2021) 179:747–53. doi: 10.1016/j.procs.2021.01.063
5. Mudele O, Frery AC. Dengue vector population forecasting using multisource earth observation products and recurrent neural networks. *IEEE J Select Topics Appl Earth Observ Remote Sens*. (2021) 14:4390–404. doi: 10.1109/JSTARS.2021.3073351
6. Teng Y, Bi D, Guo X. Predicting the epidemic potential and global diffusion of mosquito-borne diseases using machine learning. *SSRN Electr J*. (2018). doi: 10.2139/ssrn.3260785
7. Manogaran G, Lopez D. Disease surveillance system for big climate data processing and dengue transmission. *Int J Ambient Comput Intell*. (2017) 8:88–105. doi: 10.4018/IJACI.2017040106
8. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: a case study in China. *PLoS Negl Trop Dis*. (2017) 11:e0005973. doi: 10.1371/journal.pntd.0005973
9. Wu X, Lu Y, Zhou S, Chen L, Xu B. Impact of climate change on human infectious diseases: empirical evidence and human adaptation. *Environ Int*. (2016) 86:14–23. doi: 10.1016/j.envint.2015.09.007
10. Amutha D, Juliet M. *Impact of Climate Changes on Human Health in India*. (2017). doi: 10.2139/ssrn.3071055 Available online at: https://ssrn.com/abstract=3071055 (accessed October 1, 2021).
11. Sharma V, Kumar A, Panat L. Malaria outbreak prediction model using machine learning. *Int J Adv Res Comp Eng Technol*. (2015) 4:1–15. Available online at: https://www.ijert.org/malaria-outbreak-prediction-using-machine-learning
12. Lopez D, Gunasekaran M, Senthil Murugan B. Spatial big data analytics of influenza epidemic in Vellore, India. In: *IEEE International Conference on Big Data*. Vellore (2014).
13. Patz JA, Engelberg D, Last J. The effects of changing weather on public health. *Ann Rev Public Health*. (2000) 21:271–307. doi: 10.1146/annurev.publhealth.21.1.271
14. Du Y-D, Wang X-W, Yang X-F, Ma W-J. Impacts of climate change on human health and adaptation strategies in South China. *Adv Clim Change Res*. (2013) 4:208–14. doi: 10.3724/SP.J.1248.2013.208
15. Portier CJ, Thigpen TK, Carter SR, Dilworth CH. A human health perspective on climate change: a report outlining the research needs on the human health effects of climate change. *Environ Health Persp*. (2017) 621–625. doi: 10.1289/ehp.1002272
16. Mohapatra P, Tripathi NK, Pal I, Shrestha S. Determining suitable machine learning classifier technique for prediction of malaria incidents attributed to climate of Odisha. *Int J Environ Health Res*. (2021) 30:1–17. doi: 10.1080/09603123.2021.1905782
17. Sur A, Sah RP, Pandya S. Milk storage system for remote areas using solar thermal energy and adsorption cooling. *Materials Today*. (2020) 28:1764–70. doi: 10.1016/j.matpr.2020.05.170
18. Cheng J, Bambrick H, Frentiu FD, Devine G, Yakob L, Xu Z, et.al. Extreme weather events and dengue outbreaks in Guangzhou, China: a time-series quasi-binomial distributed lag non-linear model. *Int J Biometeorol*. (2021) 65:1033–42. doi: 10.1007/s00484-021-02085-1
19. Ngabo D, Wang D, Iwendi C, Anajemba JH, Ajao LA, Biamba C. Blockchain-based security mechanism for the medical data at fog computing architecture of internet of things. *Electronics*. (2021) 10:2110. doi: 10.3390/electronics10172110
20. Xu J, Xu K, Li Z, Meng F, Tu T, Xu L, et al. Forecast of dengue cases in 20 Chinese cities based on the deep learning method. *Int J Environ Res Public Health*. (2020) 17:453. doi: 10.3390/ijerph17020453
21. Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, et al. Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan, including outbreak predictions based on machine learning. *Geospatial Health*. (2019) 14:771. doi: 10.4081/gh.2019.771
22. Appice A, Gel YR, Iliev I, Lyubchich V, Malerba D. A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico. In: *IEEE Access*. vol. 8 (2020). p. 52713–25. Available online at: https://aquila.usm.edu/fac_pubs/17904 (accessed October 1, 2021).
23. Benedum CM, Shea KM, Jenkins HE, Kim LY, Markuzon N. Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS Negl Trop Dis*. (2020) 14:e0008710. doi: 10.1371/journal.pntd.0008710
24. Nkiruka O, Prasad R, Clement O. Prediction of malaria incidence using climate variability and machine learning. *Inf Med*. (2020) 22:100508. doi: 10.1016/j.imu.2020.100508
25. Stolerman LM, Maia PD, Kutz JN. Forecasting dengue fever in Brazil: an assessment of climate conditions. *PLoS ONE*. (2019) 14:e0220106. doi: 10.1371/journal.pone.0220106
26. Javed AR, Sarwar MU, Khan S, Iwendi C, Mittal M, Kumar N. Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition. *Sensors*. (2020) 20:2216. doi: 10.3390/s20082216
27. Kumar N, Susan S. COVID-19 pandemic prediction using time series forecasting models. In: *2020 11th International Conference on Computing, Communication and Networking Technologies*. Kharagpur: IEEE (2020). p. 1–7.
28. Marinucci GD. Building resilience against climate effects—a novel framework to facilitate climate readiness in public health agencies. *IJER Public Health*. (2014) 11:6433–58. doi: 10.3390/ijerph110606433
29. Baker RE, Mahmud AS, Metcalf CJE. Dynamic response of airborne infections to climate change: predictions for varicella. *Climatic Change*. (2018) 148:547–60. doi: 10.1007/s10584-018-2204-4
30. Hathaway J. Health implications of climate change: a review of the literature about the perception of the public and health professionals. *Curr Environ Health Rep*. (2018) 5:197–204. doi: 10.1007/s40572-018-0190-3
31. Baghdad M. Climate change and simulation of cardiovascular disease mortality: a case study of Mashhad, Iran. *Iran J Public Health*. (2017) 46:396–407. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5395536/

32. Pandya S, Sur A, Kotecha K. Smart epidemic tunnel: IoT-based sensor-fusion assistive technology for COVID-19 disinfection. *Int J Pervas Comput Commun.* (2020). doi: 10.1108/IJPCC-07-2020-0091
33. Garg D, Goel P, Pandya S, Ganatra A, Kotecha K. A deep learning approach for face detection using YOLO. In: *2018 IEEE Punecon.* Pune: IEEE (2018). p. 1–4.
34. Ghayvat H, Pandya S, Shah S, Mukhopadhyay SC, Yap MH, Wandra KH. Advanced AODV approach for efficient detection and mitigation of wormhole attack in MANET. In: *2016 10th International Conference on Sensing Technology (ICST).* Nanjing (2016). p. 1–6. doi: 10.1109/ICSensT.2016.7796286
35. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis.* (2016) 214:S375–9. doi: 10.1093/infdis/jiw400
36. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M. The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors.* (2020) 20:2559. doi: 10.3390/s20092559
37. Munasinghe A, Premaratne HL. Towards an early warning system to combat dengue. *Int J Comp Sci Electro Engg.* (2013) 1:252–6. Available online at: https://www.researchgate.net/profile/Noel-Fernando/publication/328189956_2nd_International_Conference_on_Computer_Electrical_Electronics_and_Mechanical_Engineering_IMCEEME'13/links/5bbd7e4292851c7fde3760a7/2nd-International-Conference-on-Computer-Electrical-Electronics-and-Mechanical-Engineering-IMCEEME13.pdf
38. Finkel AM. A "solution-focused" comparative risk assessment of conventional and synthetic biology approaches to control mosquitoes carrying the dengue fever virus. *Environ Syst Decis.* (2018) 38:177–97. doi: 10.1007/s10669-018-9688-3
39. Gomide J. Dengue surveillance based on a computational model of the spatiotemporal locality of Twitter. *Proceedings of the 3rd International Web Science Conference.* (2011). p. 1–8. doi: 10.1145/2527031.2527049
40. Iwendi C, Anajemba JH, Biamba C, Ngabo D. Security of things intrusion detection system for smart healthcare. *Electronics.* (2021) 10:1375. doi: 10.3390/electronics10121375
41. Srivastava A, Jain S, Miranda R, Patil S, Pandya S, Kotecha K. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Computer Science.* (2021) 7:e369. doi: 10.7717/peerj-cs.369
42. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. In: *IEEE Access.* vol. 8 (2020). p. 28462–74. doi: 10.1109/ACCESS.2020.2968537

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Patil and Pandya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fpubh.2021.797808

COVIDSAVIOR: A Novel Sensor-Fusion and Deep Learning Based Framework for Virus Outbreaks

Sharnil Pandya, Anirban Sur and Nitin Solke*

Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, India

The presented deep learning and sensor-fusion based assistive technology (Smart Facemask and Thermal scanning kiosk) will protect the individual using auto face-mask detection and auto thermal scanning to detect the current body temperature. Furthermore, the presented system also facilitates a variety of notifications, such as an alarm, if an individual is not wearing a mask and detects thermal temperature beyond the standard body temperature threshold, such as 98.6°F (37°C). Design/methodology/approach—The presented deep Learning and sensor-fusion-based approach can also detect an individual in with or without mask situations and provide appropriate notification to the security personnel by raising the alarm. Moreover, the smart tunnel is also equipped with a thermal sensing unit embedded with a camera, which can detect the real-time body temperature of an individual concerning the prescribed body temperature limits as prescribed by WHO reports. Findings—The investigation results validate the performance evaluation of the presented smart face-mask and thermal scanning mechanism. The presented system can also detect an outsider entering the building with or without mask condition and be aware of the security control room by raising appropriate alarms. Furthermore, the presented smart epidemic tunnel is embedded with an intelligent algorithm that can perform real-time thermal scanning of an individual and store essential information in a cloud platform, such as Google firebase. Thus, the proposed system favors society by saving time and helps in lowering the spread of coronavirus.

Keywords: deep learning, sensor-fusion, COVID-19, novel corona virus, auto-face mask detection, auto-thermal scanning, YOLOv4, body temperature

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Pooja Shah, Gandhinagar Institute of Technology, India
Abdul Rehman Javed, Air University, Pakistan
Chennareddy Vijay Simha Reddy, Middlesex University, United Kingdom

****Correspondence:***
Sharnil Pandya
sharnil.pandya@sitpune.edu.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 October 2021*
Accepted: *02 November 2021*
Published: *30 November 2021*

Citation:
Pandya S, Sur A and Solke N (2021) COVIDSAVIOR: A Novel Sensor-Fusion and Deep Learning Based Framework for Virus Outbreaks. Front. Public Health 9:797808. doi: 10.3389/fpubh.2021.797808

1. INTRODUCTION

The COVID-19 outbreak has given sleepless nights to the entire world for the last 6 months. In recent times, an economy-driven country such as India has recorded more than 1,500,000 cases (1). Considering these facts, researchers are putting enormous efforts into developing innovative solutions to deal with the current pandemic. Furthermore, according to WHO reports, the

WHO has put enormous efforts into assisting various developed and developing countries in terms of masks, ventilators, hospital beds, face shields, and other essential health equipments (2). Furthermore, the WHO has also prepared strict guidelines for lock-downs, social-distancing, and testing of COVID-19 suspects. In developing countries, such as India, along with the physical efforts of doctors, nurses, and the paramedical staff, the government has utilized digital technologies in a variety of ways, such as "Aarogya–Setu app," which can assist citizens in identifying near-by COVID-19 patients using Bluetooth and GPS based remote tracking technologies. However, looking at the impact and the spread of coronavirus, fellow researchers need to put more effort into designing and developing COVID-19 solutions (3–7). It is a fact that fellow researchers have proposed various innovative approaches to deal with the current precarious situation, including automatic sanitizer systems for the disinfection of medical equipment and individuals, thermal scanning guns, and many more. However, the development of COVID-19 assistive systems has remained an open issue for fellow researchers (8–10). In addition to this, various state governments have formulated strict health prevention policies and have made genuine efforts to disinfect various geographical regions using various sanitizers, such as disinfectants and insecticides. The government has also executed numerous health-awareness campaigns for the well-being of society. In some cases, the government has also penalized mischievous individuals who have attempted to break the health and safety procedures. It has also been recorded that the police have to patrol in the surrounding areas to ensure strict adherence to health and safety precautions, such as wearing a mask, maintaining social distancing etc. However, it is also observed and predicted by the world's health experts that society will not have access to coronavirus vaccination in the near future. So the fellow researchers need to keep on developing efficient and effective solutions to avoid unwanted circumstances (11). For this purpose, we have carried out a detailed and rigorous analysis of the existing methodologies and identified certain research gaps. It is also imperative that if the situation continues to persist for a longer time, we have to learn to live with the given problems by keeping ourselves and sound at the workplace. To resolve the discussed issues, in the undertaken study, we have presented a deep learning and sensor-fusion-based approach for detecting a face mask and real-time body temperature for COVID-19, which is term as "Smart Facemask and Thermal scanning kiosk" throughout this paper. The presented deep learning and sensor-fusion-based approach is designed to detect an individual with or without mask conditions and notify security personnel by raising the alarm. Moreover, the smart tunnel is also equipped with a thermal sensing unit embedded with a camera, which can detect the real-time body temperature of an individual concerning the prescribed body temperature limits as prescribed by WHO reports. The presented article is organized as follows: Section 2 discusses the state of the art methodologies. Section 3 discusses the necessity of the presented system, section 4 and 5 discusses the design and experimental setup, architecture design, sensing arrangements, deployments, and the detailed workflow of the presented system. Finally, section 6 provides the concluding remarks.

2. RELATED WORK

Yang and his team have proposed an object detection auto-masking neural network for capturing discriminative objects. In this study, a variety of simulation-based study has been carried out. The proposed system is not tested in a real-time environment (12). Joshi R. has presented a health kiosk model using Computational Fluid Dynamics (CFD) simulation-based studies. However, the presented approach has been tested under the simulation environment only (13). Fan and his fellow researchers have proposed an auto-lung segmentation system to detect CT-scans of coronavirus patients. However, the proposed system did not facilitate the scanning of face-masks and the body temperature of humans (14). Maurya and his team have presented an innovative method for disinfecting humans. However, the proposed system did not facilitate any kind of face-mask and body temperature scanning (15). Ghayvat and his team have presented a system to keep the track of social-distancing policies between two persons. For this purpose, the proposed system has used smart cities, such as ITS infrastructures. However, the proposed system did not discuss any kind of scanning approach for the human body and face-masks (2). Abbas and his fellow mates have presented a chest X-ray scanning system of COVID-19 suspects. However, the presented system is not capable to do thermal scanning of a human body or identifying obscured faces (16). Apostolopoulos and Mpesiana have presented a chest X-ray scanning system of COVID-19 suspects. However, the presented systems were not capable to do thermal scanning of a human body or identifying obscured faces (17). Poon and his research team have proposed a system that is capable to carry out obstetric and gynecological scans. The main purpose of the presented system was to do scanning of medical equipment (18). Kim and Lee have presented a respiratory illness scanning system to detect diseases, such as COPD. However, the proposed system did not facilitate thermal scanning and the detection of obscured faces (19). Ucar and fellow researchers have presented an AI-based CT-scan scanning system to detect coronavirus symptoms. However, the presented system did not discuss anything like detecting body temperature or face-masks (20). Kwon and fellow researchers have presented a human scanning system to detect COVID-19 like symptoms for a drive-through. However, the presented system did not discuss anything like detecting body temperature or face-masks (21). Majid et al. (22) researchers have presented a sensor-fusion based wearable device that can scan and disinfect human hands from coronavirus infections (22). Takagi and Yagishita have presented a detailed comparison and discussion of various health and safety policies related to COVID-19 issues. However, the system did not propose any system for COVID-19 like diseases (23). Farman and team have presented a geofencing based real-time tracking system for COVID-19 patients. However, the presented system did not discuss any scanning mechanisms for the current pandemic situations. However, the system was not

designed for the disinfection of healthy patients (24). Lippi and fellow researchers have analyzed and discussed numerous bio-safety policies and precautions. However, the presented system did not discuss any scanning approaches for humans (25). Mahammedi et al. (26) have presented a novel solar power-based system to disinfect things such as mobile, key, wallet, and many more. However, the system did not possess the capability to perform any kind of scanning. Pandya et al. (27) have presented a novel face-detection system that can detect a human with fully or partially covered faces. However, the system did not possess the capability to detect face-masks. Ghayvat et al. (28) have presented a sensor-fusion based smart aging system to monitor the elderly well beings. However, the system did not facilitate the scanning of faces and body temperature of a human.

3. NECESSITY OF A SMART FACE MASK AND THERMAL SCANNING KIOSK

Recent fellow researchers have made efforts to implement a cost-effective pandemic tunnel to disinfect humans from COVID-19 like infections. In this study, an IoT based sensor fusion assistive tunnel has been presented, which can disinfect an individual from the possibility of coronavirus infections. The major contributions of the presented system are as follows: a deep learning and IoT based sensor fusion assistive framework has been proposed to do real-time detection of individuals with face-mask and without face-mask conditions. In the presented research work, a solar-powered smart tunnel has been presented, installed and deployed at the entrance of the Symbiosis institute of technology, Pune. The contributions of the proposed work are as follows:

1. The presented system can also detect an outsider who is entering the building with or without mask conditions using You Only Look Once 4 (YOLOv4) computer vision algorithm (29) and notifies the security control room by raising appropriate alarms.
2. The presented smart face-mask detection and thermal scanning kiosk can function using solar energy during the day, and it functions using a solar power bank at night time. This functionality has been provided by light-dependent resistor (LDR) sensing unit placed in a tunnel.
3. The smart tunnel also provides the facility to do real-time attendance of all the staff entering the smart tunnel, along with thermal scanning and face-mask detection.
4. In the end, web and mobile interface has been designed to provide daily, weekly and monthly reports of the recorded body temperatures of individuals, along with in-out timestamp values. The investigation results validate the performance evaluation of the presented smart face-mask and thermal scanning kiosk. The presented smart tunnel is embedded with an intelligent algorithm that can perform real-time thermal scanning of an individual and stores the essential information in a cloud platform, such as google firebase.

4. DESIGN AND EXPERIMENTAL SETUP

In the experimental setup, each sensor is placed in different positions in the house, with a single ESP8266 attached to it, where the basic data processing takes place. ESP8266 comes with a Wi-Fi module attached to it, using which the processed data is then transferred logged into the Raspberry Pi B+ server. To avoid data duplication of erroneous data, pre-processing of sensor data is done at the ESP8266 level, thus ensuring

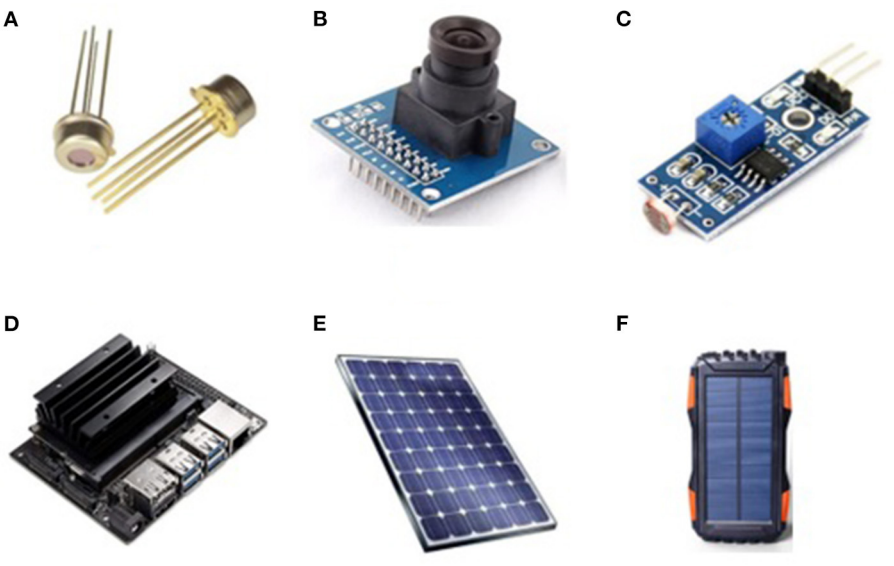


FIGURE 1 | Design and experimental setup of a smart epidemic tunnel **(A)** thermopile temperature sensor a Robocraze OV7670 **(B)** 300KP VGA Camera **(C)** an LDR sensing unit **(D)** a NVIDIA Jetson Nano **(E)** a solar cell **(F)** a solar power bank.

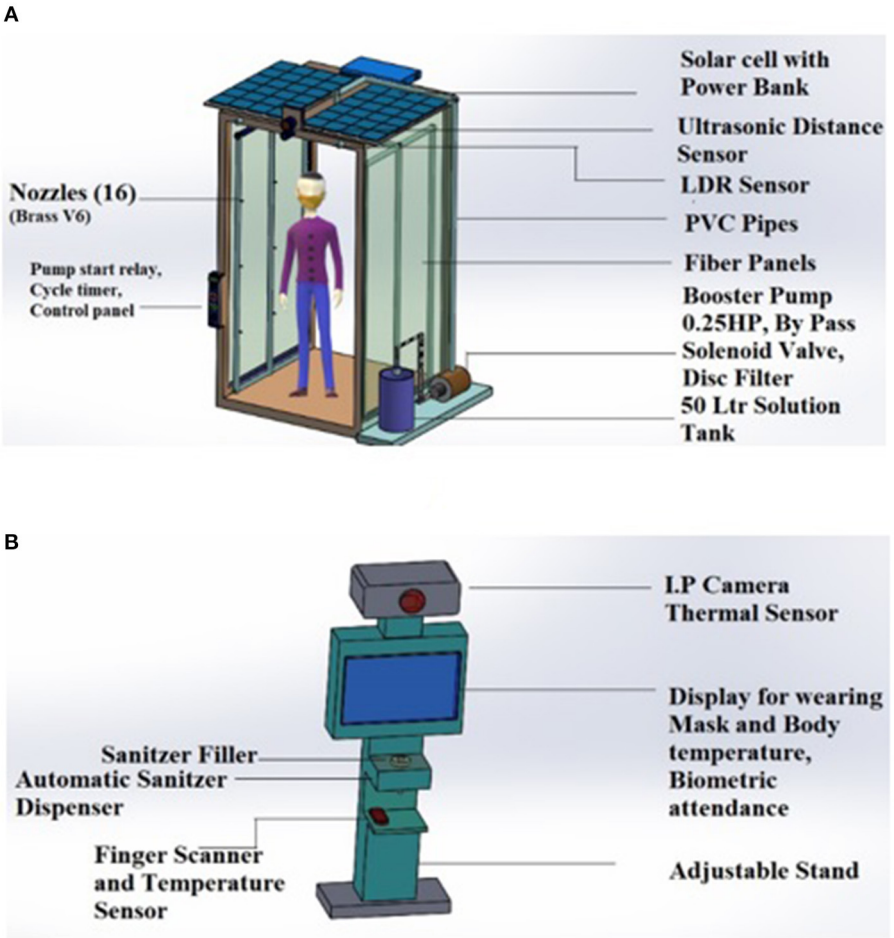


FIGURE 2 | Prototype design of a **(A)** smart epidemic tunnel **(B)** face-mask and thermal scanning Kiosk.

that Raspberry pi is only used as the local server where clean data is stored in the whole setup. Logged Data is periodically sent to the Cloud storage to ensure backup of data is kept in case of any system failure, along with the logs of whole systems working, which can also be later on utilized for systems debugging. The presented smart tunnel is an extended version of the smart epidemic tunnel, which had facilitated the ultrasonic sensor-based sanitizer system and timestamp-based notifications. In the presented research work, we have proposed a customized smart face-mask and thermal scanning kiosk equipped in a previously presented smart epidemic tunnel. The presented system is flexible enough to be placed outside public spots such as malls, university buildings, companies, bus-stop, hospital ICU units, vegetable markets, railway stations, and many more. **Figures 1A–F** represent a design and experimental setup used in the conducted experiments such as a Thermopylae sensor, a VGA camera, LDR sensing unit, Netson NVIDIA controller, a solar cell, and a solar power bank. **Figure 2** represents a prototype design of the presented smart face mask and thermal scanning kiosk equipped in a smart epidemic tunnel.

5. METHODOLOGY

Fellow researchers have made efforts to implement a cost-effective pandemic tunnel to disinfect humans from COVID-19 like infections. In this study, an IoT based sensor fusion assistive tunnel has been presented, which can disinfect an individual from the possibility of coronavirus infections. The major contributions of the presented PWP system are as follows: (i) a deep learning and IoT based sensor fusion assistive framework have been proposed to do real-time detection of individuals with face-mask and without face-mask conditions. In the presented research work, a solar-powered smart tunnel has been presented, installed, and deployed at the entrance of the Symbiosis institute of technology, Pune. (i) The presented system can also detect an outsider who is entering the building with or without mask conditions using YOLOv4 computer vision algorithm and notifies the security control room by raising appropriate alarms. (ii) The presented smart face-mask detection and thermal scanning kiosk can function using solar energy during the day, and it functions using a solar power bank at night. This functionality has been provided by an (LDR) sensing unit

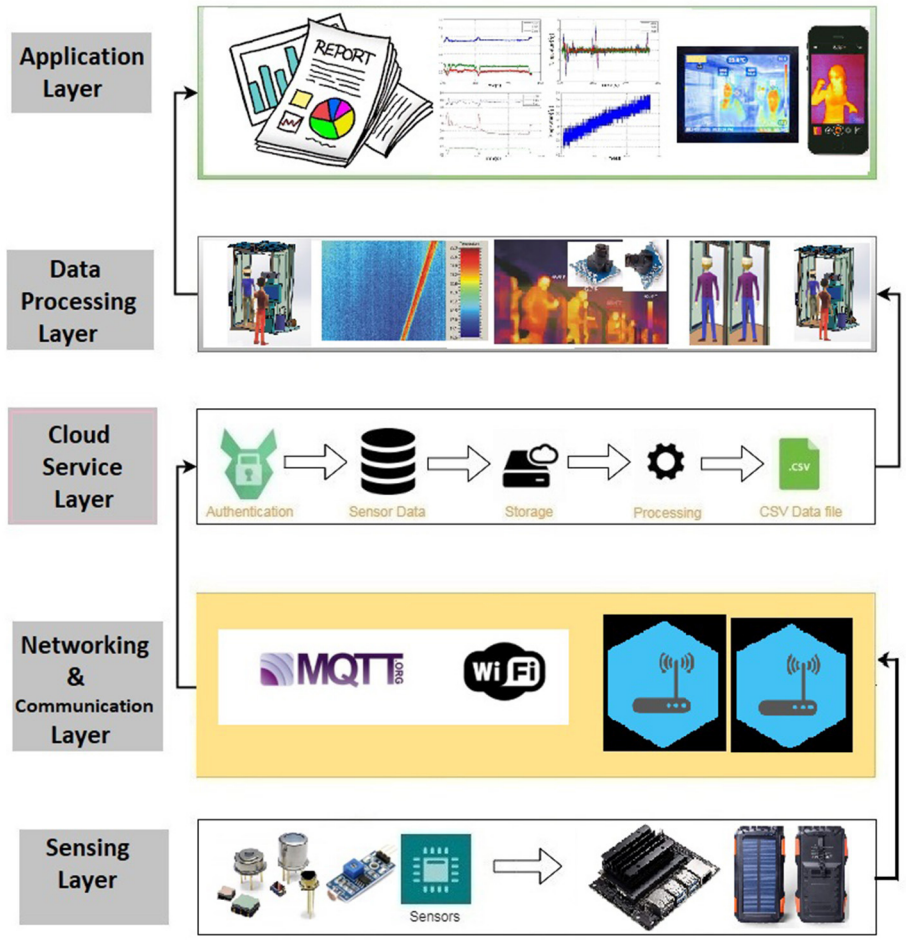


FIGURE 3 | An architectural design of a smart facemask and thermal scanning system prototype design of a facemask and thermal scanning Kiosk.

placed in a tunnel. (iii) The smart tunnel also provides the facility to do real-time attendance of all the staff entering the smart tunnel, along with thermal scanning and face-mask detection. (iv) In the end, web and mobile interface has been designed to provide daily, weekly, and monthly reports of the recorded body temperatures of individuals, along with in-out timestamp values. The investigation results validate the performance evaluation of the presented smart face-mask and thermal scanning kiosk. The presented smart tunnel is embedded with an intelligent algorithm that can perform real-time thermal scanning of an individual and stored the essential information in a cloud platform, such as google firebase. In the pandemic situation, a smart face-mask and thermal scanning kiosk have been presented, equipped in a smart epidemic tunnel. The presented research work is an extended version of the smart epidemic tunnel. In the presented research work, we have added certain novel features to enhance the presented system's performance. The presented deep learning and sensor-fusion based assistive technology (Smart Facemask and Thermal scanning kiosk) will protect the individual using auto face-mask detection and auto thermal scanning to detect the current body temperature. The presented smart tunnel is embedded with an intelligent algorithm that can perform real-time thermal scanning of an individual and stored the essential information in a cloud platform, such as google firebase. Furthermore, in the conducted experiments, the YOLOv4 computer vision algorithm has been applied to detect individuals with face-mask and without faced-mask conditions. Furthermore, the presented system also facilitates a variety of notifications such as an alarm if an individual is not wearing a mask and detects thermal temperature beyond the standard body temperature threshold such as 98.6°F (37°C). Furthermore, a solar power bank is also used for storing solar energy, which the proposed system will utilize at night. In the end, the investigation results validate the performance evaluation of the presented smart face-mask and thermal scanning kiosk.

5.1. The Layered Design of a Smart Face-Mask and Thermal Scanning Kiosk

Figure 3 represents an architectural design of the presented novel deep Learning and sensor-fusion-based approach for detecting

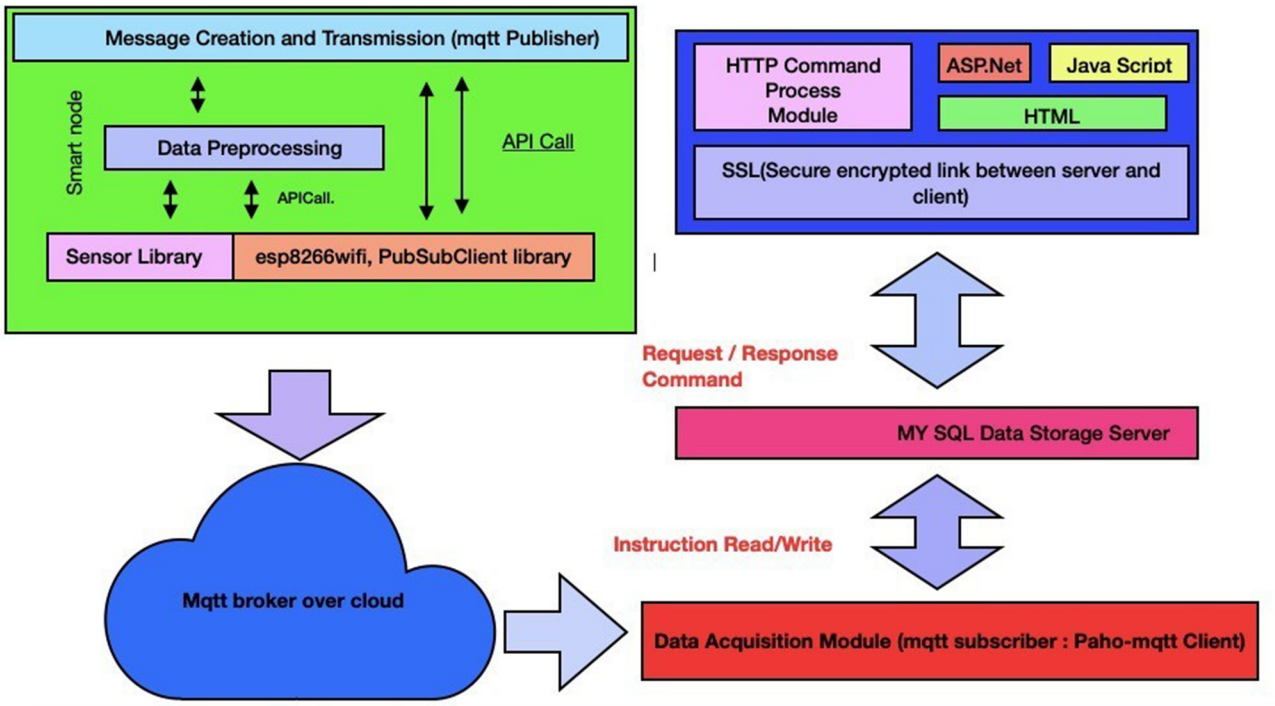


FIGURE 4 | A back-end design of a smart face-mask thermal scanning system.

a face mask and real-time body temperature for COVID-19 outbreak situations. The extended version of the presented has been placed at the Symbiosis Institute of Technology, Pune, which has been implemented for performing auto-face mask detection and thermal scanning of various Symbiosis stakeholders such as faculty members, students, and external visitors. The presented architectural design contains five layers: (i) physical (sensing) layer, (ii) communication and networking layer, (iii) cloud computing (google firebase storage) layer, (iv) data processing layer, and (v) application layer. The main purpose of the physical sensing layer is to detect the body temperature of individuals who are present in the projection areas of a VGA camera equipped in the presented thermal scanning kiosk. This sensing layer contains various sensing units such as the LDR sensing unit and a thermopile temperature sensor. Furthermore, this layer also contains an intelligent NVIDIA Netson controller, along with a solar cell and a solar power bank. A YOLOv4 computer vision algorithm has been installed in an NVIDIA Netson controller to detect humans with face masks and without face-mask conditions. In addition to this, this layer also facilitates a solar power backup using a solar cell during the day-time and a solar power bank during the night time. Again, in the day-time, the presented smart face-mask and thermal scanning kiosk start functioning using a solar cell.

5.1.1. Sensing Layer

The main purpose of the physical sensing layer is to detect the body temperature of individuals present in the projection areas of a VGA camera equipped in the presented thermal scanning kiosk. This sensing layer contains various sensing units, such as the LDR sensing unit and a thermopile temperature sensor. Furthermore, this layer also contains an intelligent NVIDIA Netson controller, along with a solar cell and a solar power bank. A YOLOv4 computer vision algorithm has been installed in an NVIDIA Netson controller to detect humans with face masks and without face-mask conditions. In addition to this, this layer also facilitates a solar power backup using a solar cell during daytime and a solar power bank during nighttime. Again, in the daytime, the presented smart face-mask and thermal scanning kiosk start functioning using a solar cell.

5.1.2. Communication and Networking Layer

The networking and communication layer is responsible for interfacing and transmitting information between a physical(sensing) layer, cloud services layer, and an application layer using a broker architecture of an MQTT protocol. Furthermore, it transmits scanned body temperature values, timestamps, and the count of people wearing or not wearing a face mask on a google firebase platform. **Figure 4** represents a back-end design of a smart epidemic tunnel.

5.1.3. Communication and Networking Layer

The networking and communication layer is responsible for interfacing and transmitting information between a physical(sensing) layer, cloud services layer, and an application layer using a broker architecture of an MQTT protocol.

Furthermore, it transmits scanned body temperature values, timestamps, and the count of people wearing or not wearing a face mask on a google firebase platform. **Figure 4** represents a back-end design of a smart epidemic tunnel.

5.1.4. Cloud Computing (Google Firebase) Layer

The cloud computing layer facilitates the storage of the temperature values of individuals scanned using a thermal sensing unit with timestamp values and the count of people with face-mask and without face-mask variations. Furthermore, it keeps track of various individuals who have accessed the smart face-mask and thermal scanning kiosk during the day or night timings. In the conducted experiments, the Google Firebase cloud computing platform has been used for storage purposes.

5.1.5. Processing Layer

The processing layer fetches the temperature and face-mask related values from the Google Firebase database *via* MQTT broker architecture. This layer performs two essential operations: (i) the processing layer detects the body temperature of individuals using a thermal sensor placed in a smart thermal scanning kiosk (ii) it also facilitates two essential bifurcation of the people who are wearing or not wearing face-masks using a YOLOv4 computer vision algorithm deployed in an NVIDIA Netson controller. Eventually, the processed information will be passed on to the application layer for generating various reports.

5.1.6. Application Layer

The application layer received organized information from the processing layer and represented it in various graphical and tabular representations. Furthermore, the application layer also consists of the web and mobile interface, which provides daily, weekly, and monthly updates such as timestamp-based temperature detection reports, the count of people wearing or not wearing face-masks, and the number of individuals who have access to the presented face-mask and thermal scanning kiosk during the day and night timings.

6. THE DETAIL WORKING OF THE FACE-MASK AND THERMAL SCANNING KIOSK

As shown in **Figure 5**, the presented smart system performs two key operations during this phase: (i) it also facilitates the bifurcation of the people based on a variety of face mask conditions such as with and without face mask conditions using a YOLOv4 computer vision algorithm deployed in an NVIDIA Netson controller. (ii) The presented smart system can detect the body temperature of individuals present in the projection areas of a Robocraze OV7670 300KP VGA camera using a thermopile temperature sensing unit.

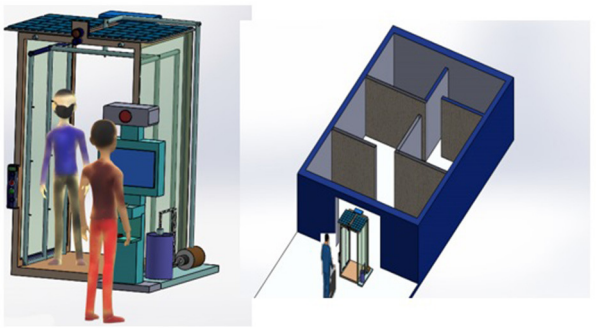


FIGURE 6 | A prototype design of a face-mask detection system using a VGA camera and YOLOv4 deep learning methodology.

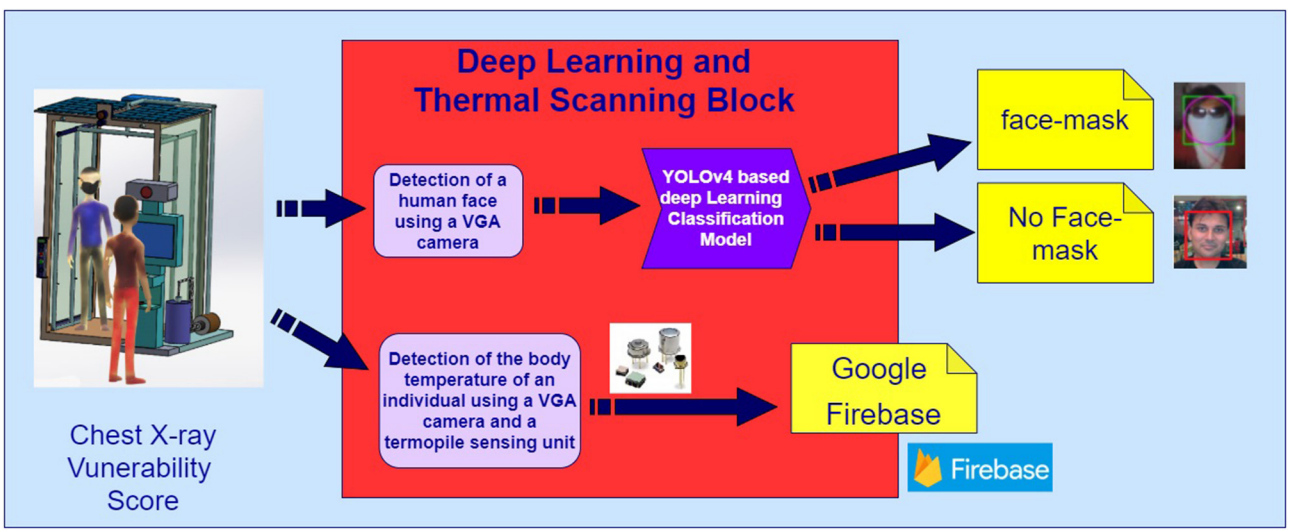


FIGURE 5 | A topology design of a smart facemask and thermal scanning system.

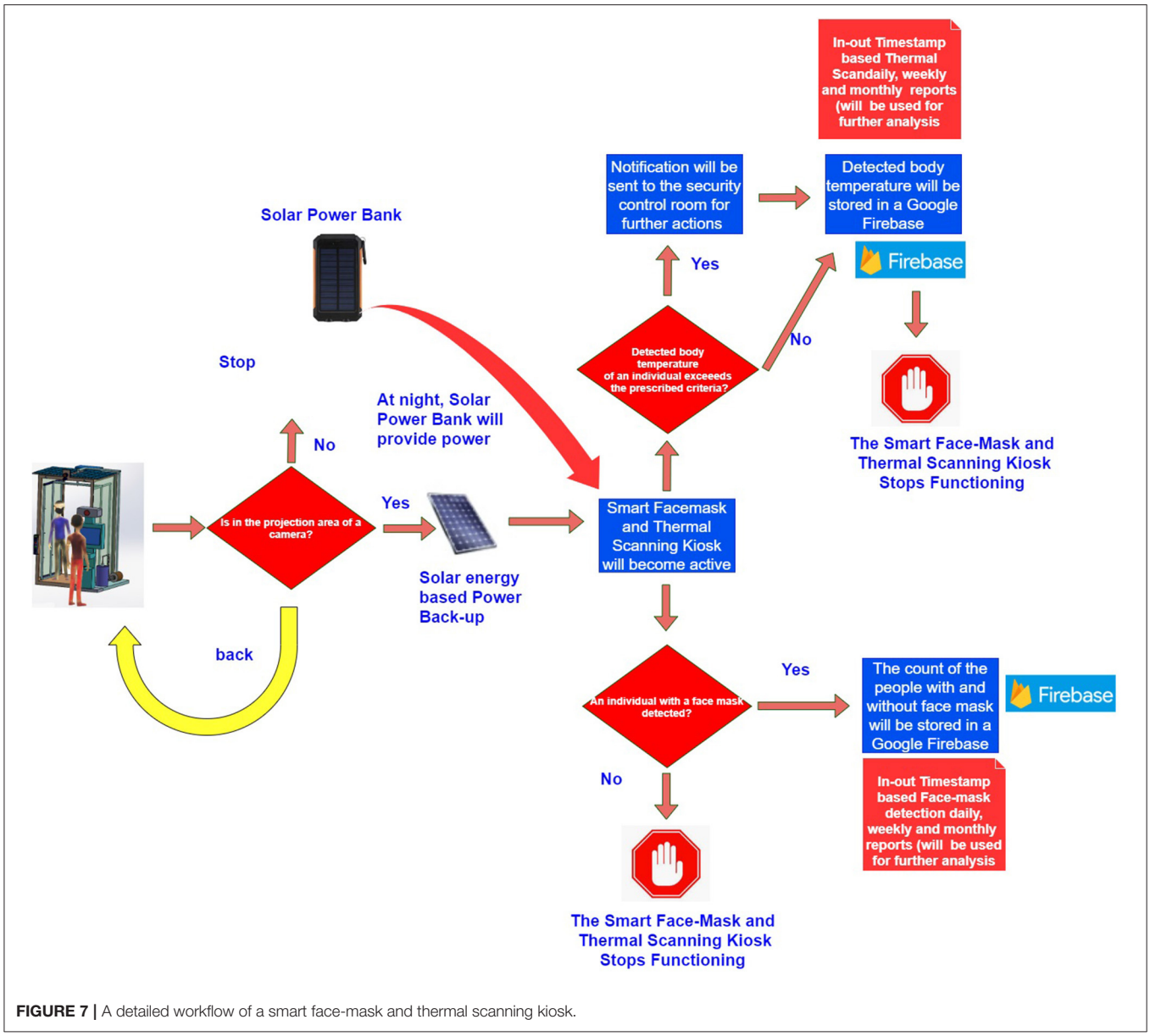


FIGURE 7 | A detailed workflow of a smart face-mask and thermal scanning kiosk.

6.1. The Detailed Working of the Face-Mask Detection Methodology

The face-mask detection feature becomes active when an individual comes into the projection area of a VGA camera, as shown in **Figure 6**. In the conducted experiments, we have used a customized dataset of 1,000 face samples. After an individual is detected by a VGA camera placed in a smart kiosk, a YOLOv4 computer vision algorithm deployed in an NVIDIA Netson controller process the captured frame; during this phase, a YOLOv4 algorithm makes the use of cross-stage partial connection and separates the feature maps in two different parts: (i) the first part of the feature map bypasses the dense block and directly moves toward the transition block. (ii) The second part went through the dense block and reached the transition block. This kind of design increases the efficiency of the YOLOv4 algorithm by reducing the complexity of the presented approach. In addition to this, the YOLOV4 algorithm also utilizes cross-stage-partial connections with CSPDarknet-53 feature extraction methodology, which enables the presented algorithm to achieve very high objection detection accuracy. **Figure 7** represents a YOLOv4 methodology, and **Figure 8** represents a detailed workflow of the presented smart face-detection and thermal scanning kiosk. As shown in **Figure 8**, after an individual is detected by a VGA camera placed in a smart kiosk using a YOLOv4 computer vision algorithm, it will process the captured frame of an individual to identify whether the captured individual is wearing a face-mask to protect against COVID-19 or not. Then, the algorithm will transmit the processed information to

```
yer     filters    size              input                output
  0 conv     32  3 x 3 / 1   608 x 608 x   3   ->   608 x 608 x  32  0.639 BFLOPs
  1 conv     64  3 x 3 / 2   608 x 608 x  32   ->   304 x 304 x  64  3.407 BFLOPs
  2 conv     32  1 x 1 / 1   304 x 304 x  64   ->   304 x 304 x  32  0.379 BFLOPs
  3 conv     64  3 x 3 / 1   304 x 304 x  32   ->   304 x 304 x  64  3.407 BFLOPs
  4 res    1                 304 x 304 x  64   ->   304 x 304 x  64
  5 conv    128  3 x 3 / 2   304 x 304 x  64   ->   152 x 152 x 128  3.407 BFLOPs
  6 conv     64  1 x 1 / 1   152 x 152 x 128   ->   152 x 152 x  64  0.379 BFLOPs
  7 conv    128  3 x 3 / 1   152 x 152 x  64   ->   152 x 152 x 128  3.407 BFLOPs
  8 res    5                 152 x 152 x 128   ->   152 x 152 x 128
  9 conv     64  1 x 1 / 1   152 x 152 x 128   ->   152 x 152 x  64  0.379 BFLOPs
 10 conv    128  3 x 3 / 1   152 x 152 x  64   ->   152 x 152 x 128  3.407 BFLOPs
 11 res    8                 152 x 152 x 128   ->   152 x 152 x 128
 12 conv    256  3 x 3 / 2   152 x 152 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 13 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 14 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 15 res   12                  76 x  76 x 256   ->    76 x  76 x 256
 16 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 17 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 18 res   15                  76 x  76 x 256   ->    76 x  76 x 256
 19 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 20 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 21 res   18                  76 x  76 x 256   ->    76 x  76 x 256
 22 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 23 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 24 res   21                  76 x  76 x 256   ->    76 x  76 x 256
 25 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 26 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 27 res   24                  76 x  76 x 256   ->    76 x  76 x 256
 28 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 29 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 30 res   27                  76 x  76 x 256   ->    76 x  76 x 256
 31 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 32 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 33 res   30                  76 x  76 x 256   ->    76 x  76 x 256
 34 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
 35 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
 36 res   33                  76 x  76 x 256   ->    76 x  76 x 256
 37 conv    512  3 x 3 / 2    76 x  76 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 38 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 39 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 40 res   37                  38 x  38 x 512   ->    38 x  38 x 512
 41 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 42 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 43 res   40                  38 x  38 x 512   ->    38 x  38 x 512
 44 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 45 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 46 res   43                  38 x  38 x 512   ->    38 x  38 x 512
 47 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 48 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 49 res   46                  38 x  38 x 512   ->    38 x  38 x 512
 50 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 51 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 52 res   49                  38 x  38 x 512   ->    38 x  38 x 512
 53 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 54 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 55 res   52                  38 x  38 x 512   ->    38 x  38 x 512
 56 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 57 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 58 res   55                  38 x  38 x 512   ->    38 x  38 x 512
 59 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 60 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 61 res   58                  38 x  38 x 512   ->    38 x  38 x 512
 62 conv   1024  3 x 3 / 2    38 x  38 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 63 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 64 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 65 res   62                  19 x  19 x1024   ->    19 x  19 x1024
 66 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 67 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 68 res   65                  19 x  19 x1024   ->    19 x  19 x1024
 69 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 70 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 71 res   68                  19 x  19 x1024   ->    19 x  19 x1024
 72 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 73 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 74 res   71                  19 x  19 x1024   ->    19 x  19 x1024
 75 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 76 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 77 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 78 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 79 conv    512  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 512  0.379 BFLOPs
 80 conv   1024  3 x 3 / 1    19 x  19 x 512   ->    19 x  19 x1024  3.407 BFLOPs
 81 conv    255  1 x 1 / 1    19 x  19 x1024   ->    19 x  19 x 255  0.189 BFLOPs
 82 yolo
 83 route  79
 84 conv    256  1 x 1 / 1    19 x  19 x 512   ->    19 x  19 x 256  0.095 BFLOPs
 85 upsample           2x     19 x  19 x 256   ->    38 x  38 x 256
 86 route  85 61
 87 conv    256  1 x 1 / 1    38 x  38 x 768   ->    38 x  38 x 256  0.568 BFLOPs
 88 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 89 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 90 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 91 conv    256  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 256  0.379 BFLOPs
 92 conv    512  3 x 3 / 1    38 x  38 x 256   ->    38 x  38 x 512  3.407 BFLOPs
 93 conv    255  1 x 1 / 1    38 x  38 x 512   ->    38 x  38 x 255  0.377 BFLOPs
 94 yolo
 95 route  91
 96 conv    128  1 x 1 / 1    38 x  38 x 256   ->    38 x  38 x 128  0.095 BFLOPs
 97 upsample           2x     38 x  38 x 128   ->    76 x  76 x 128
 98 route  97 36
 99 conv    128  1 x 1 / 1    76 x  76 x 384   ->    76 x  76 x 128  0.568 BFLOPs
100 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
101 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
102 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
103 conv    128  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 128  0.379 BFLOPs
104 conv    256  3 x 3 / 1    76 x  76 x 128   ->    76 x  76 x 256  3.407 BFLOPs
105 conv    255  1 x 1 / 1    76 x  76 x 256   ->    76 x  76 x 255  0.754 BFLOPs
106 yolo
Loading weights from yolov4.weights...Done!
data/dog.jpg: Predicted in 0.461468 seconds.
 training accuracy: 97.17%  training loss: 3.88%
 testing accuracy: 96.12%   testing loss:2.83
```

FIGURE 8 | An implementation snapshot of a YOLOv4 deep learning algorithm.

the Google Firebase and generate various timestamp-based daily, weekly, and monthly reports in both scenarios.

6.2. The Detailed Working of the Thermal-Scanning Methodology

The thermal scanning feature of a presented smart kiosk starts functioning when an individual is detected in the projection area of a VGA camera. Once an individual is detected, the presented system will scan their body temperature using a thermal sensor and compare it with the standard body temperature threshold such as 98.6°F (37°C). If the presented kiosk records body temperature of an individual beyond the prescribed limits, it will immediately notify the security control room for further action. Otherwise, timestamp-based records of scanned body temperatures will be stored in Google Firebase for further analysis in both scenarios. The YOLOV4 methodology is divided into two parts: (i) the first part bypasses the dense block and directly moves toward the transition block. (ii) The second part goes through the dense block and reaches the transition block. This kind of design increases the efficiency of the YOLOv4 algorithm by reducing the complexity of the presented approach. In addition to this, the YOLOV4 algorithm also utilizes cross-stage-partial connections with CSPDarknet-53 feature extraction methodology, which enables the presented algorithm to achieve very high objection detection accuracy. **Figure 7** represents a YOLOv4 methodology, and **Figure 8** represents a detailed workflow of the presented smart face-detection and thermal scanning kiosk.

7. RESULTS AND DISCUSSIONS

In this section, a detailed discussion of the obtained results of the presented smart face-mask and thermal scanning kiosk for two different approaches (i) the YOLOv4 based face-mask detection methodology (ii) thermopile temperature sensor VGA camera-based body temperature detection methodology.

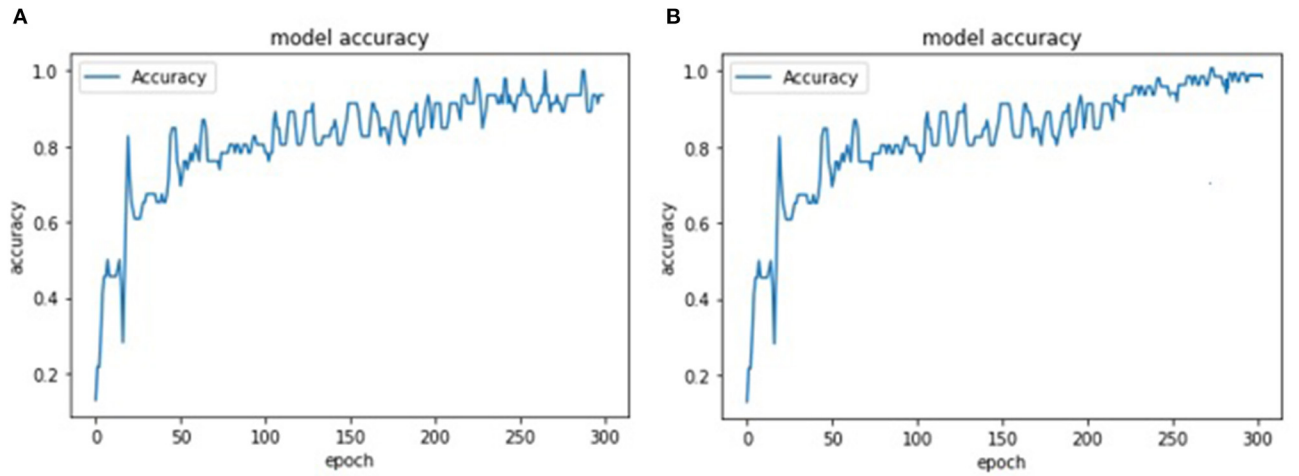


FIGURE 9 | A comparison of a YOLOv4 training and testing accuracy metrics **(A)** training accuracy **(B)** testing accuracy.

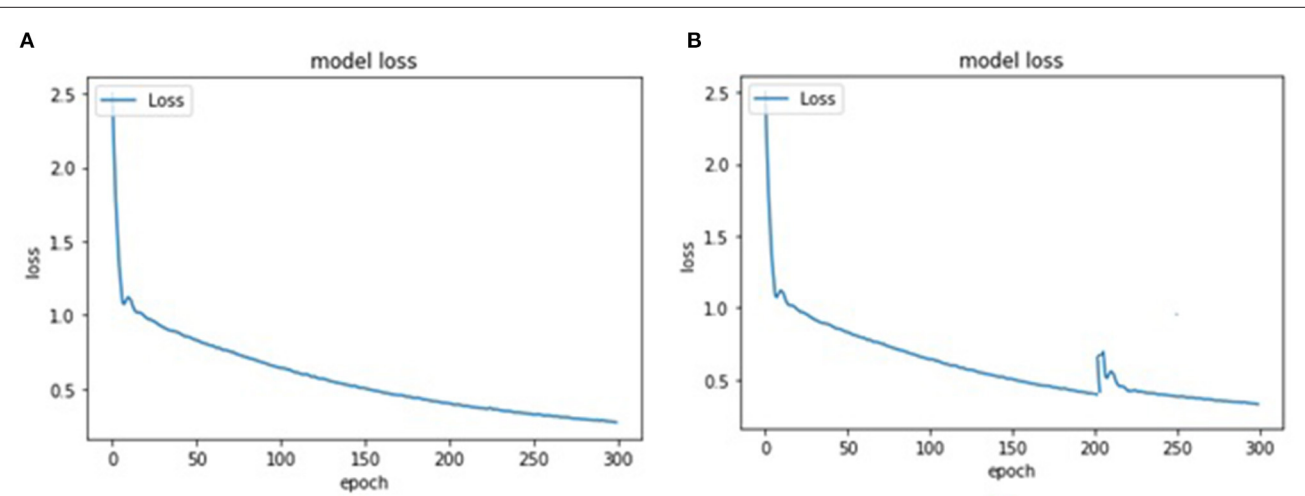


FIGURE 10 | A comparison of a YOLOv4 training and testing accuracy metrics **(A)** training loss **(B)** testing loss.

7.1. Results Discussion of the YOLOv4 Based Face-Mask Detection Methodology

Figure 9 represents training and testing snapshots of the presented YOLOv4 deep learning methodology. A customized dataset of 1,000 face images was used in the conducted experiments and trained using the YOLOv4 methodology. As described in section 5.2.1, in the conducted experiments, we have used CSPDarknet-53 feature extraction methodology along with the presented YOLOv4 deep learning approach. As shown in **Figure 9**, we have modified the presented YOLOv4 algorithm to achieve high accuracy in detecting individuals with and without face-mask conditions. Based on the conducted experiments, the presented YOLOV4 methodology has recorded face-mask detection training and testing accuracy of 97.17 and 96.12%, as shown in **Figure 10**. As shown in **Figure 11**, based on the conducted experiments, the presented YOLOv4 deep learning algorithm has recorded very minimal training and testing loss, around 1.83 and 2.88%, respectively. **Figure 12** depicts face-mask detection samples detected using the YOLOv4 algorithm.

7.2. Results Discussion of the Thermal Scanning Based Body Temperature Detection Methodology

In the conducted experiments, we have performed real-time testing of more than 800 individuals to validate the performance of the presented body temperature detection methodology. The presented methodology has utilized a Robocraze OV7670 300KP VGA camera and thermopile temperature sensor to identify individuals exceeding the prescribed standard threshold such as 98.6°F (37°C). **Figure 13** represents the real-time thermal scanning results of more than 800 individuals concerning the recorded body temperature. Furthermore, it also depicts the

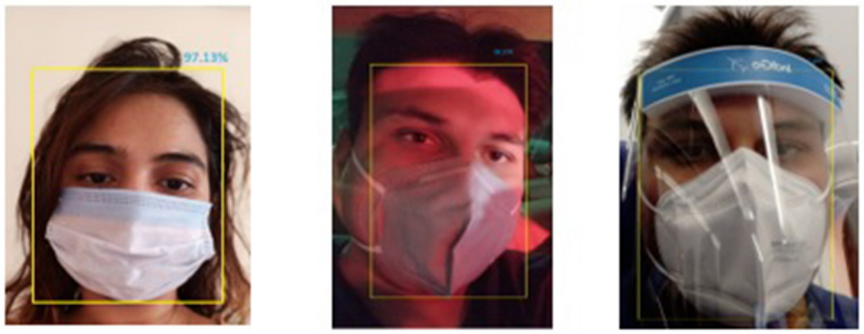


FIGURE 11 | A testing sample of YOLOv4 face-mask detection.

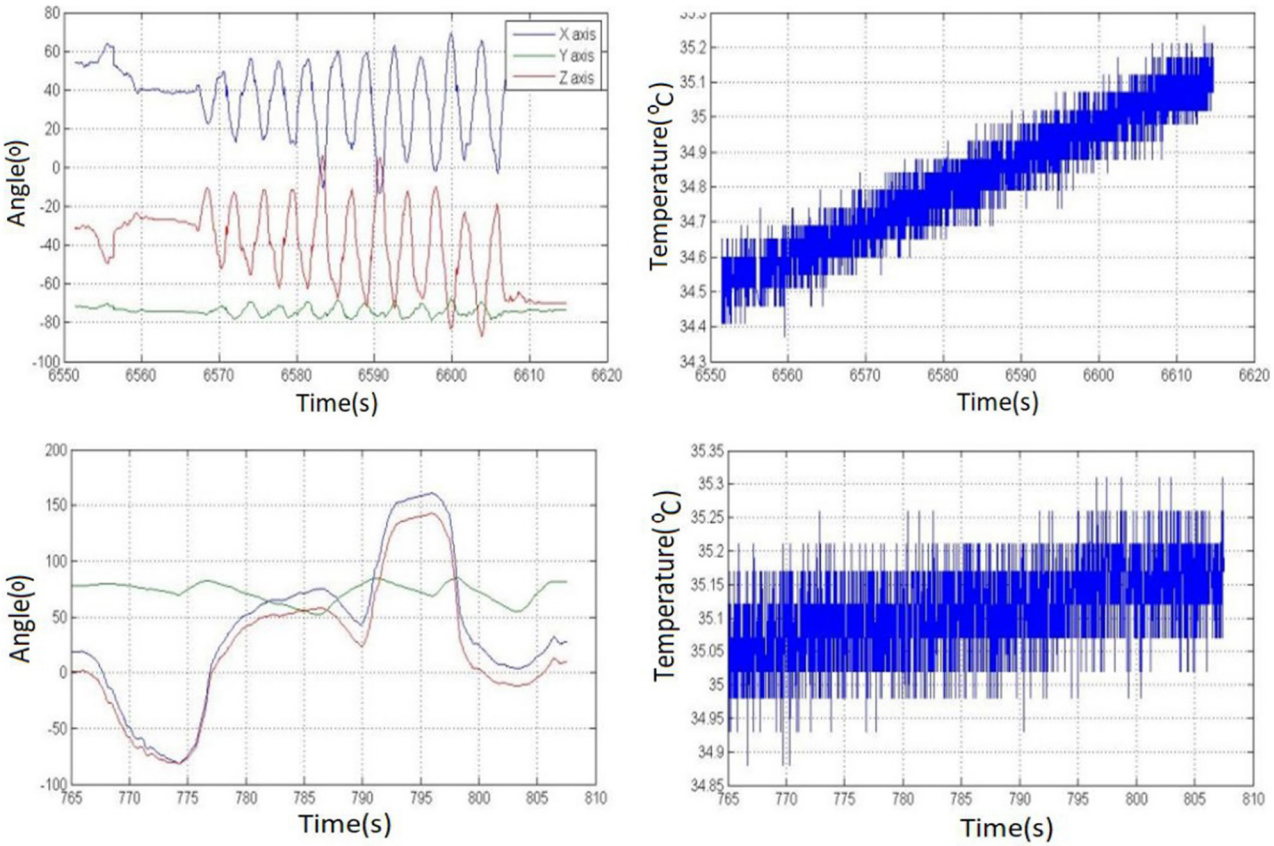


FIGURE 12 | Real-time thermal scanning results of more than 800 individuals.

graphical representation of the number of individuals who a VGA camera has captured from various angles. The presented thermal scanning mechanism has recorded a validation accuracy of 92.01% thermal scanning kiosk. Furthermore, the presented interfaces present the daily, weekly, and monthly reports of the counts of individuals, along with in-out timestamps fetched from the google firebase cloud computing platform. Furthermore, it also keeps track of the consumed power and also provides information on power usage reports.

7.3. The Web and Mobile Interface of a Smart Face-Mask and Thermal Scanning Kiosk

Figure 13 represents a GUI-based web and mobile interface design of the presented smart face-mask and thermal scanning kiosk. The presented interfaces present the daily, weekly, and monthly reports of the counts of individuals, along with in-out timestamps fetched from the google firebase cloud computing

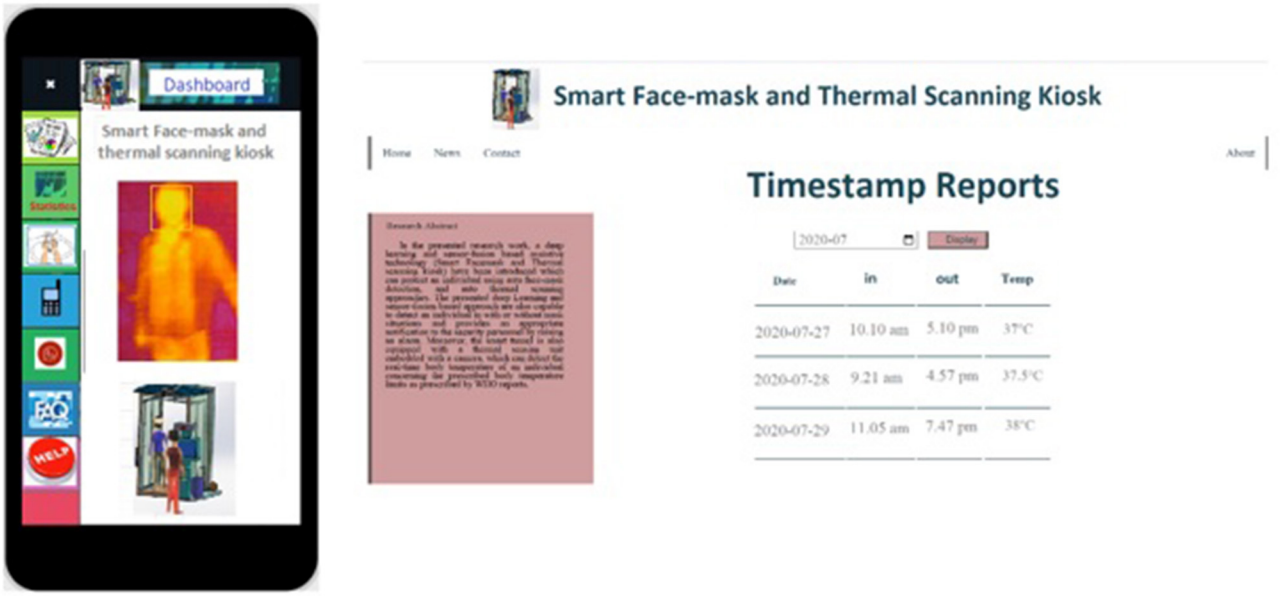


FIGURE 13 | The web and mobile interface of a smart face-mask and thermal scanning kiosk.

platform. Furthermore, it also keeps track of the consumed power and also provides information on power usage reports.

8. CONCLUSION AND FUTURE ENHANCEMENTS

This study presents a deep learning and sensor-fusion-based assistive technology (Smart Facemask and Thermal scanning kiosk) to detect individuals with and without mask conditions. The presented research work is an extended version of the smart epidemic tunnel presented previously. In the undertaken study, the presented smart face-mask and thermal scanning kiosk have facilitated two essential tasks (i) the bifurcation of the people based on a variety of face mask conditions such as with and without face mask conditions using a YOLOv4 computer vision algorithm deployed in an NVIDIA Netson controller. In addition, (ii) it can also detect the body temperature of individuals present in the projection areas of a Robocraze OV7670 300KP VGA camera using a Thermopylae temperature sensing unit. The investigation results validate the performance evaluation of the presented smart face-mask and thermal scanning mechanism. The major findings of this study are as follows: (i) The presented system can detect an outsider who is entering the building with or without mask conditions and notifies the security control room by raising appropriate alarms. (ii) The presented smart face-mask detection and thermal scanning kiosk can function using solar energy during the day, and it functions using a solar power bank at night. This functionality has been provided by an LDR sensing unit placed in a tunnel. (iii) The presented smart kiosk also provides the facility to do real-time timestamp-based attendance of all the staff entering the smart tunnel, along with thermal scanning of the body temperature and face-mask detection. Based on the experiments, the presented YOLOV4 methodology recorded face-mask detection training and testing accuracy of 97.17 and 96.12%. Furthermore, the presented thermal scanning mechanism has recorded a validation accuracy of 92.01%. (iv) In the end, web and mobile interface has been designed to provide daily, weekly, and monthly reports of the recorded body temperatures of individuals, along with in-out timestamp values. In the future, the proposed research work can be extended with various scenarios, such as obscured faces, half, or fully obscured human faces with plastic, leather, or any other materials.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SP and AS: conceptualization and data collection and interpretation and data curation. SP, AS, and NS: manuscript editing. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of COVID-19 outbreak in Pakistan. *Comput Mater Continua*. (2021) 69:1253–69. doi: 10.32604/cmc.2021.015655
2. Ghayvat H, Awais M, Gope P, Pandya S, Majumdar S. Recognizing suspect and predicting the spread of contagion based on mobile phone location data (counteract): a system of identifying covid-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence. *Sus Cities Soc*. (2021) 69:102798. doi: 10.1016/j.scs.2021.102798
3. Pandya S, Ghayvat H. Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence. *Adv Eng Inform*. (2021) 47:101238. doi: 10.1016/j.aei.2020.101238
4. Aslam B, Javed AR, Chakraborty C, Nebhen J, Raqib S, Rizwan M. Blockchain and ANFIS empowered IoMT application for privacy preserved contact tracing in COVID-19 pandemic. *Pers Ubiquitous Comput*. (2021) 22:1–17. doi: 10.1007/s00779-021-01596-3
5. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sus Cities Soc*. (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
6. Garg D, Goel P, Pandya S, Ganatra A, Kotecha K. A deep learning approach for face detection using YOLO. In *2018 IEEE Punecon*. Pune: IEEE (2018). p. 1–4.
7. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, et al. CP-BDHCA: blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform*. (2021) doi: 10.1109/JBHI.2021.3097237
8. Pandya S, Sur A, Kotecha K. Smart epidemic tunnel: IoT-based sensor-fusion assistive technology for COVID-19 disinfection. *Int J Pervasive Comput Commun*. (2020) doi: 10.1108/IJPCC-07-2020-0091
9. Sur A, Sah RP, Pandya S. Milk storage system for remote areas using solar thermal energy and adsorption cooling. *Mater Today Proc*. (2020) 28:1764–70. doi: 10.1016/j.matpr.2020.05.170
10. Srivastava A, Jain S, Miranda R, Patil S, Pandya S, Kotecha K. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Comput Sci*. (2021) 7:e369. doi: 10.7717/peerj-cs.369
11. Ghayvat H, Pandya S, Shah S, Mukhopadhyay S, Yap M, Wandra K. Advanced AODV approach for efficient detection and mitigation of wormhole attack in MANET. In: 2016 *10th International Conference on Sensing Technology (ICST)*. Nanjing: IEEE (2016). p. 1–6.
12. Yang L, Liu J, Tang X. Object detection and viewpoint estimation with auto-masking neural network. In: *European Conference on Computer Vision*. Glasgow: Springer (2014). p. 441–55.
13. Joshi JR. COVSACK: an innovative portable isolated and safe COVID-19 sample collection kiosk with automatic disinfection. *Trans Indian Natl Acad Eng*. (2020) 5:269–75. doi: 10.1007/s41403-020-00139-1
14. Fan DP, Zhou T, Ji GP, Zhou Y, Chen G, Fu H, et al. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Trans Med Imaging*. (2020) 39:2626–37. doi: 10.1109/TMI.2020.2996645
15. Maurya D, Gohil MK, Sonawane U, Kumar D, Awasthi A, Prajapati AK, et al. Development of autonomous advanced disinfection tunnel to tackle external surface disinfection of COVID-19 virus in public places. *Trans Indian Natl Acad Eng*. (2020) 5:281–7. doi: 10.1007/s41403-020-00141-7
16. Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl Intell*. (2021) 51:854–64. doi: 10.1007/s10489-020-01829-7
17. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med*. (2020) 43:635–40. doi: 10.1007/s13246-020-00865-4
18. Poon L, Abramowicz J, Dall'Asta A, Sande R, Ter Haar G, Maršal K, et al. ISUOG safety committee position statement on safe performance of obstetric and gynecological scans and equipment cleaning in context of COVID-19. *Ultrasound ObstetrGynecol*. (2020) 55:709–12. doi: 10.1002/uog.22027
19. Kim SI, Lee JY. Walk-through screening center for COVID-19: an accessible and efficient screening system in a pandemic situation. *J Korean Med Sci*. (2020) 35:e154. doi: 10.3346/jkms.2020.35.e154
20. Ucar F, Korkmaz D. COVIDiagnosis-Net: deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses*. (2020) 140:109761. doi: 10.1016/j.mehy.2020.109761
21. Kwon KT, Ko JH, Shin H, Sung M, Kim JY. Drive-through screening center for COVID-19: a safe and efficient screening system against massive community outbreak. *J Korean Med Sci*. (2020) 35:e123. doi: 10.3346/jkms.2020.35.e123
22. El Majid B, Motahhir S, El Hammoumi A, Lebbadi A, El Ghzizal A. Preliminary design of a smart wristband disinfectant to help in covid-19 fight. *Inventions*. (2020) 5:32. doi: 10.3390/inventions5030032
23. Takagi G, Yagishita K. Principles of disinfectant use and safety operation in medical facilities during coronavirus disease 2019. (COVID-19) outbreak. *SN Comprehens Clin Med*. (2020) 2:1041–4. doi: 10.1007/s42399-020-00413-x
24. Ullah F, Haq HU, Khan J, Safeer AA, Asif U, Lee S. Wearable IoTs and geo-fencing based framework for COVID-19 remote patient health monitoring and quarantine management to control the pandemic. *Electronics*. (2021) 10:2035. doi: 10.3390/electronics10162035
25. Lippi G, Adeli K, Ferrari M, Horvath AR, Koch D, Sethi S, et al. Biosafety measures for preventing infection from COVID-19 in clinical laboratories: IFCC taskforce recommendations. *Clin Chem Lab Med*. (2020) 58:1053–62. doi: 10.1515/cclm-2020-0633
26. Mahammedi NA, Mahammedi A. Standalone solar-powered ultraviolet mobile disinfectant: bringing solar energy in the global fight against COVID-19. *Preprints*. (2020) 2020:2020050415. doi: 10.20944/preprints202005.0415.v1
27. Pandya S, Ghayvat H, Kotecha K, Awais M, Akbarzadeh S, Gope P, et al. Smart home anti-theft system: a novel approach for near real-time monitoring and smart home security for wellness protocol. *Appl Syst Innov*. (2018) 1:42. doi: 10.3390/asi1040042
28. Ghayvat H, Awais M, Pandya S, Ren H, Akbarzadeh S, Chandra Mukhopadhyay S, et al. Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. *Sensors*. (2019) 19:766. doi: 10.3390/s19040766
29. Sengupta S. *YOLO Application in COVID-19 Use Cases*. (2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Pandya, Sur and Solke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fpubh.2021.759032

A Comprehensive Review on the Optical Micro-Electromechanical Sensors for the Biomedical Application

Anup M. Upadhyaya [1,2,3], Mohammad Kamrul Hasan [4], S. Abdel-Khalek [5], Rosilah Hassan [4]*, Maneesh C. Srivastava [1,2], Preeta Sharan [3], Shayla Islam [6]*, Asma Mohammed Elbashir Saad [7] and Nguyen Vo [8]*

[1] Department of Mechanical Engineering, Amity School of Engineering and Technology (ASET), Amity University, Noida, Lucknow, India, [2] Department of Mechanical Engineering, The Oxford College of Engineering, Bangalore, India, [3] Department of Electronics and Communication Engineering, The Oxford College of Engineering, Bangalore, India, [4] Network and Communication Technology Lab, Center for Cyber Security, Faculty of Information Science and Technology, The National University of Malaysia (UKM), Bangi, Malaysia, [5] Department of Mathematics and Statistics, College of Science, Taif University, Taif, Saudi Arabia, [6] Institute of Computer Science and Digital Innovation, University College Sedaya International (UCSI) University, Kuala Lumpur, Malaysia, [7] Department of Physics College of Science and Humanities in AL-Kharj, Prince Sattam Bin Abdulaziz University, AL-Kharj, Saudi Arabia, [8] Department of Information Technology, Victorian Institute of Technology, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Ciro Rodriguez,
National University of San Marcos, Peru
Manpreet Manna,
Sant Longowal Institute of Engineering and Technology, India

***Correspondence:**
Mohammad Kamrul Hasan
hasan.iium@gmail.com;
mkhasan@ukm.edu.my
Rosilah Hassan
roislah@ukm.edu.my
Shayla Islam
shayla@ucsiuniversity.edu.my

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 15 August 2021
Accepted: 15 September 2021
Published: 02 December 2021

Citation:
Upadhyaya AM, Hasan MK, Abdel-Khalek S, Hassan R, Srivastava MC, Sharan P, Islam S, Saad AME and Vo N (2021) A Comprehensive Review on the Optical Micro-Electromechanical Sensors for the Biomedical Application. Front. Public Health 9:759032. doi: 10.3389/fpubh.2021.759032

This study presented an overview of current developments in optical micro-electromechanical systems in biomedical applications. Optical micro-electromechanical system (MEMS) is a particular class of MEMS technology. It combines micro-optics, mechanical elements, and electronics, called the micro-opto electromechanical system (MOEMS). Optical MEMS comprises sensing and influencing optical signals on micron-level by incorporating mechanical, electrical, and optical systems. Optical MEMS devices are widely used in inertial navigation, accelerometers, gyroscope application, and many industrial and biomedical applications. Due to its miniaturised size, insensitivity to electromagnetic interference, affordability, and lightweight characteristic, it can be easily integrated into the human body with a suitable design. This study presented a comprehensive review of 140 research articles published on photonic MEMS in biomedical applications that used the qualitative method to find the recent advancement, challenges, and issues. The paper also identified the critical success factors applied to design the optimum photonic MEMS devices in biomedical applications. With the systematic literature review approach, the results showed that the key design factors could significantly impact design, application, and future scope of work. The literature of this paper suggested that due to the flexibility, accuracy, design factors efficiency of the Fibre Bragg Grating (FBG) sensors, the demand has been increasing for various photonic devices. Except for FBG sensing devices, other sensing systems such as optical ring resonator, Mach-Zehnder interferometer (MZI), and photonic crystals are used, which still show experimental stages in the application of biosensing. Due to the requirement of sophisticated fabrication facilities and integrated systems, it is a tough choice to consider the other photonic system. Miniaturisation of complete FBG device for biomedical applications is the future scope of work.

Even though there is a lot of experimental work considered with an FBG sensing system, commercialisation of the final FBG device for a specific application has not been seen noticeable progress in the past.

Keywords: optical MEMS (optical micro electro mechanical system), microcantilever, fibre bragg grating, pressure sensor, intraocular pressure, blood pressure, urodynamic, orthodontic

INTRODUCTION

Micro-opto electromechanical system or optical micro-electromechanical system combines optical components with micro electromechanical system technology. The micro-level mechanical elements with micro-level motion manipulate the optical signals by actuation provided by electronic structures in a system called micro-opto electromechanical system (MOEMS). Many people are working on optical micro-electromechanical system (MEMS) technology to develop sensitive sensors. Electronic devices size varies from millimetre, micrometre, to nanometre, but mechanical element size ranges from four to five times the electronic devices. This requires sophisticated fabrication technology for optical MEMS. To fabricate the optical MEMS sensor, the technology involved are bulk micromachining, deep x-ray lithography, and surface micromachining. Microfabrication consists of, the process of generating micro ma-chines. Functionality testing of optical and electronic structures is the first phase of MOEMS device fabrication. In this context, the development and analysis of a mechanical element is the second phase of MOEMS device advancement. The third phase involves material deposition, etching, and pattern. The optical design process consists of choosing suitable geometry and assessing optical. Electromechanical design involves various mechanical factors such as stiffness, material parameter, electrostatic, and thermal. MOEMS device design depends on different parameters. Different factors that need to be considered during the design of optical MEMS are mechanical, electrical, thermal, ecological, biological, optical, and chemical aspects based on different applications. The shape and dimension considered during the design of optical MEMS play a significant role in the proper working of optical MEMS devices. Stress distribution due to different geometrical aspects and optimization is a factor that influences the MOEMS design. **Figure 1A** shows the MOEMS overview chart (1).

Strength, stiffness, hardness, wear resistance, abrasion, creep, and fatigue resistance of the material used during the design are parameters evident in the sound optical MEMS system. The refractive index, an optical property, is the governing factor of any MOEMS device. In the optical design, we know that if the refractive index changes directly affect the performance of any optical MEMS structures, then the particular assignment of refractive index is a must before analysis. The optical MEMS sensor consists of two main parts: (1) the micro sensing element and transduction unit, which powers the supply unit with optical media, and (2) the micro sensing element, which senses the input signals and converts them into output in the required form of the transduction unit. Different types of optical MEMS sensors that are possible in a real scenario are displacement, temperature, bio-sensor, chemical, optical, and pressure sensors. The optical MEMS system can also act as a microactuator by actuating micro actuating elements (2). **Figure 1B** shows the different types of optical MEMS systems.

This analysis aimed to research the literature on photonic MEMS in biomedical systems. We investigated different studies providing different backgrounds and their relationships. The following research questions (RQ) are specified. RQ1: "What are the major issues regarding using photonic MEMS?" RQ2: "What key success factors are applied to methods that influence performance enhancement?" RQ3: "What are the research gap and scope to design the optimum photonic MEMS devices in the biomedical application?"

MATERIALS AND METHODS

This research relied on a systematic review of the MOEMS and its application in the biomedical field. Parallel understanding of available research associated with optical MEMS, such as the basic principle of optical MEMS, different design configurations involved, and its future scope of improvements, were discussed. There are three key phases in this review paper. First, the principle of MOEMS and the different types of optical MEMS. Second, the specific application of optical MEMS. Third, the challenges towards optical MEMS with the current market trend and future scopes. Research papers associated with the basics of optical biosensors, integrated optical sensors, and optical MEMS for biomedical applications, and articles explaining the current status of optical MEMS sensors in the market were selected to help answer the research questions.

Different methods are employed in the design of optical MEMS. It involves modelling MOEMS as the sensor, modelling MOEMS as the actuator, modelling different substrates, beam propagation techniques, computational opto electro mechanics, and high fidelity modelling of the electromagnetic field.

High fidelity modelling of MEMS based on vertical cavity surface emitting method uses Laguerre Gaussian equation. This involves single, double, and multimode equations in consideration.

$$E(r,z) = E_o \frac{w_o}{w(z)} \frac{w_o}{w(z)} \exp\left(\frac{-r^2}{w^2(z)}\right) \frac{-r^2}{w^2(z)}$$
$$\exp\left(-ikz - ik\frac{r^2}{2R(z)}\frac{r^2}{2R(z)} + i\zeta(z)\right.$$

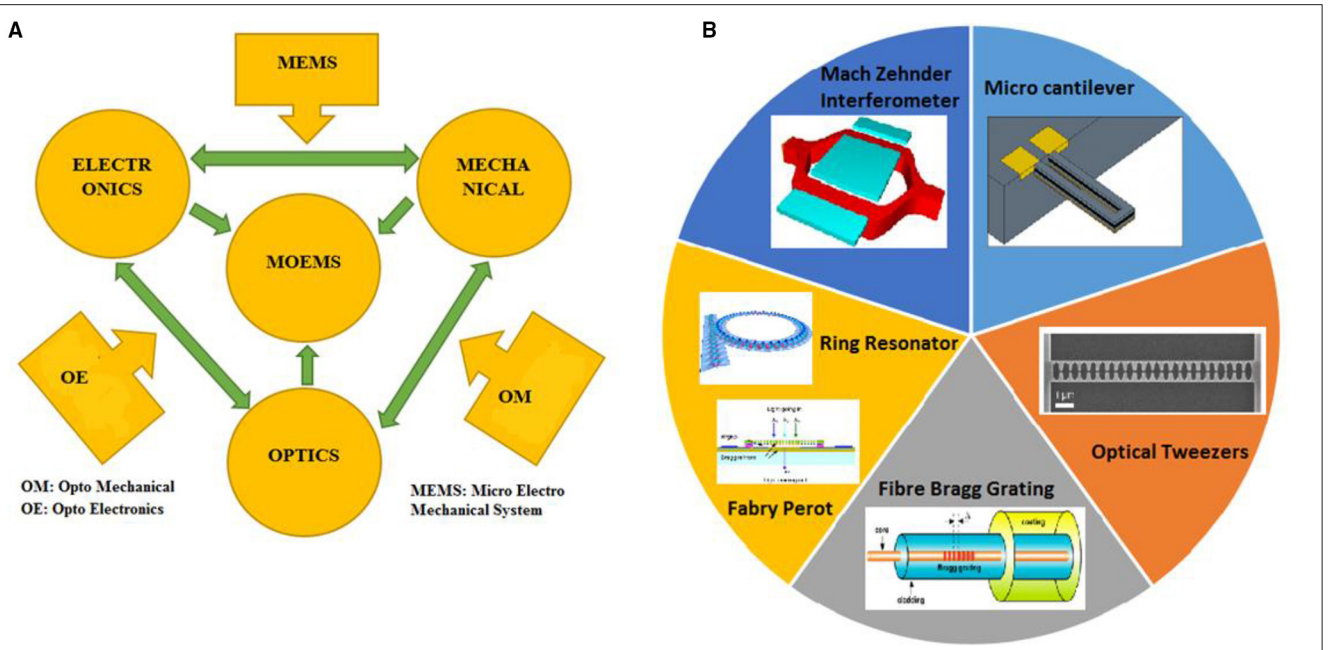


FIGURE 1 | (A) MOEMS overview chart, **(B)** different types of Optical MEMS.

Maxwell equations are used at the beginning of every optic problem.

$$\text{i.e., div D} = \rho, (2)\text{div B} = 0, (3)\text{curl E} = \frac{\mathbf{dD}}{\mathbf{dt}} + \mathbf{J}\frac{\mathbf{dD}}{\mathbf{dt}} + \mathbf{J}$$

where ρ is charge density, J is current density, E is the electric field, B is the magnetic field, and D and H are field quantities that are proportional to E and B, respectively. The last technique involved considering in the design and modelling of optical MEMS is the beam propagation technique. This technique uses Helmholtz time-dependent diffusion equation.

$$(-\nabla^2 - \frac{\omega^2}{\mathbf{v^2}}\frac{\omega^2}{\mathbf{v^2}})\, \text{q(x)}$$

Many optical MEMS devices incorporate different types of materials for sensing or actuating sensing layers. Different types of micro sensing or actuating elements are thermal forces, electrostatic forces, piezoelectric materials, and shape memory that alloys along with optical MEMS integration in a different environment. Photonic sensing layers silica, optical fibre made up of silicon, are the materials, considered in the optical sensing system. Germanium and gallium arsenide are also used in other conditions like thin films. Along with the other materials, the sensing system also consists of polymer materials. Photonic MEMS as actuator involves movable and fixed sensing elements, cold or hot arm in terms of thermomechanical actuation. **Figures 2A,B** show different sensing systems with MOEMS.

Major types of optical MEMS such as FBG sensor, optical microcantilever sensor, Mach-Zehnder interferometer (MZI),

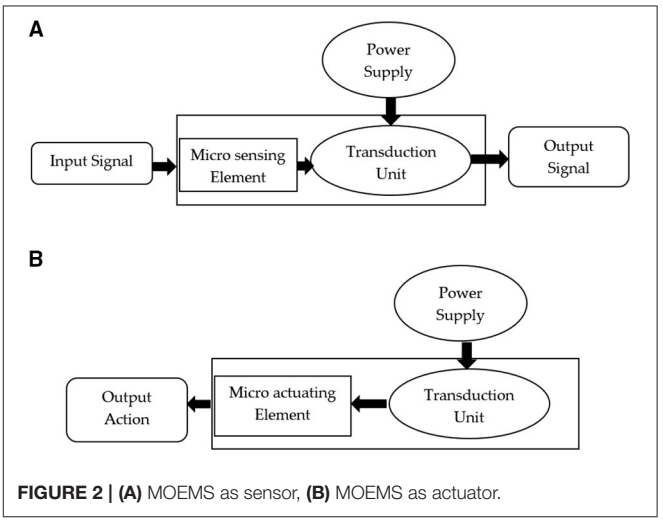


FIGURE 2 | (A) MOEMS as sensor, **(B)** MOEMS as actuator.

and photonic MEMS with microfluidic channels, as well as their applications, were discussed here. It was categorised based on its approach to integration with MEMS application and principle of operation, among others. Most of the works of literature in optical MEMS-based microcantilever sensors incorporated with photonic crystal sensing layer, ring resonator structures, and optical readout techniques were discussed. MZI is integrated with the photonic sensing layer or diaphragm for sensing the deflection in the sensing arm. The level of maturity of FBG techniques has reached to commercial stage since they are widely explored in different biomedical applications. Recent

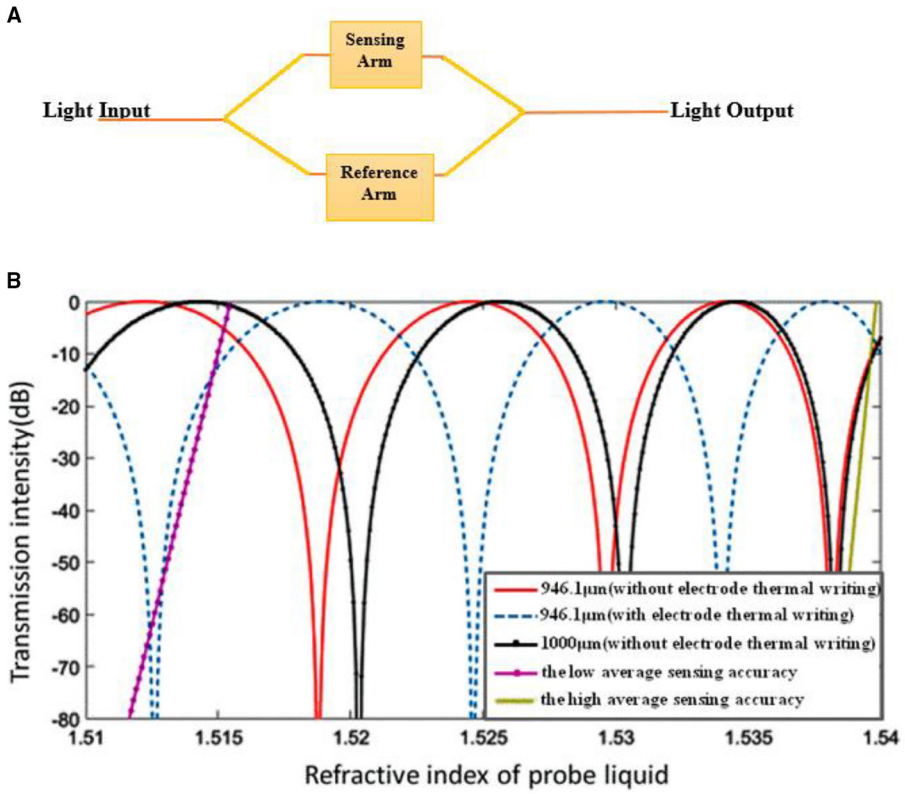


FIGURE 3 | (A) Mach-Zehnder Interferometer working process, **(B)** intensity of sensor with various length and refractive index of core (6).

advancements in photonic MEMS with microfluidic channels and its design and fabrication were also discussed in this study.

TYPE OF OPTICAL MEMS SENSORS WITH APPLICATION

There are various types of optical MEMS sensors. Depending on the application, different sensing mechanisms had been reported and presented in this manuscript. This section explained the basic principle of all possible optical sensing schemes used in optical MEMS.

Types of Optical MEMS Sensors

Optical MEMS Based Mach-Zehnder Interferometer

Photonic devices generally have extreme sensitivity to the environment and manufacturing, having wide applications. It is a device used to measure the change in phase shift between two beams. In this context, MZI with high sensitivity has three core fibres proposed by Ding et.al. The MZI is considered an optical MEMS sensor that was bringing properties such as elongation or shortening of the core made possible by applying force. Overall bending sensitivity of 15.35 and 3.11 nm/m was obtained (3). MZI is also used to propose the potential measurement of the trajectories of joint angle. Two core trapezoid cone structures have been used for the simultaneous measurement of pressure and temperature. Developed high sensitive MZI for this purpose was sensitive to laboratory seismic noise. A possible micromechanical structure with electro-optical function has been investigated with an In-P material with an integrated optical stress sensor to reduce the seismic noise (4). In another work, Jindal and Raghuwanshi discussed the photoelastic effect induced phase shift during the transmission of light in MZI. Due to maximum pressure at the centre of the diaphragm, the centre portion attached to the MZI gets deflected and brings out the shift in phase and intensity variations. Theoretical analysis of micromachined differential pressure sensor clamped with circular diaphragm and FBG sensor embedded in MZI has been analysed, wherein the sensitivity of this pressure sensor was about 0.2 pm/pa (5).

The Mach-Zehnder Interferometer shown in **Figure 3A** consists of a light input and output and a sensing and reference arm. Left Y-junction which have light input splits the light beam called as a beam splitter and right junction combines the light and called as beam combiner. Two straight waveguides between two Y-junctions locate the sensing arm at one end and the reference arm at the other ends. Output unit, the right end of Y-junction, gives the phase shift, i.e., $\Delta\Phi$. Polarised light from laser source passed to two arms of MZI, through input waveguide. Transmission of light through the sensing arm reacts with a biochemical molecule and the overall refractive index changes with the refractive index of the reference arm. Intensity modulation is obtained in the output due to variation in the

overall refractive index of a sample in the sensing arm relating to a reference arm with a normal refractive index of the medium. If $\Delta\Phi$ is the shift in phase, λ is a wavelength, and d is the length of the sensing arm, the effective refractive index is denoted by n_{eff} (7). Equations 1, 2 indicates that MZI sensor output intensity is equal to the oscillating function of interferometer phase change difference.

$$Iout = I_{sen} + I_{ref} + 2\sqrt{I_{sen}I_{ref}}(\Delta\varphi + \Delta\varphi 0) \quad (1)$$

where I_{sen} and I_{ref} are the intensity of light passing through the sensing arm and reference arm.

$$\Delta\Phi = 2\pi/\lambda.\Delta n_{eff}\text{sensing arm} \quad (2)$$

The sensitivity of the MZI is denoted by $\Delta O/\Delta n$, where ΔO is output power change and Δn changes in refractive index. The phase shift is obtained by Equation 2. In another work, PB Patel and co-authors evaluated a novel optical MZI (OMZI) with rib waveguides. The complete structure has been designed with a silicon rib waveguide. The rib width of the OMZI structure was optimised, and the effect of the rib width on wave propagation was monitored. Change in resonant peak wavelength for refractive index variation due to rib width optimizations was also analysed. It was observed from the results that mono mode behaviour was developed with the rib width of 4 μm and height of 5 μm. The rib waveguide was analysed for wavelength variation of 630 to 1,310 nm (8). In this context, Lin considered a novel electrode temperature writing waveguide based on changing refractive index for the applied heat. Different mode conditions were considered in this sensor design (6). Transmission intensity for the various refractive index of probe liquid is presented in **Figure 3B**.

In another work, Yadav et al. discussed optical waveguide-based biosensors on the subwavelength grating which were investigated with a single TE mode in 1,400 to 1,500 nm. The interferometer bend radius of 50 μm was maintained. Two-dimensional FEM simulation was performed with the Comsol Multiphysics software tool. The reference arm was considered with silica material and sensing arm incorporated with air material. There were 2-fold increases in sensitivity after including the sub-wavelength range (9) compared with the MZI proposed with single-mode fibre. Single-mode fibre was coated with methylcellulose. It has shown a humidity sensing range of 45–8% RH. A sensitivity of 0.094% RH was obtained during the simulation. An increase or decrease in the relative humidity has shown the difference of intensities up to 0.03 dB (10). Highly sensitive silicon MZI based ultrasound sensor was discussed by Ouyang et al. In this sensor, one end of the sensing arm was integrated with a thin membrane and another end of the arm was used as a reference. The ultrasound wave was excited on the sensing layer of the interferometer. The low detection limit of 0.38m Pa/Hz was obtained. In preliminary experimental conditions, the image wire phantom was formed. The properties and sensing characteristics of the sensor showed promising application such as ultrasound imaging (11). In this context, high temperature-sensitive MZI was designed with a system of optical fibre offset splicing techniques. The designed sensor has shown sensitivities for external refractive index up to −17,335 nm/RIU and temperature for 32 pm/C. The sensor was constructed with a quartz capillary sandwiched process. The sensitivity of the sensor achieved was 21.2 nm/RIU. The authors concluded its application in many biomedical and industrial concepts (12). It is observed from **Figure 4** that sensitivity changing for change in the refractive index.

In another work, Luan et al. discussed a highly sensitive lab on-chip optical ring resonator. Different types of optical ring resonator structures were considered, such as doped silicon-based dual ring resonator system and phase-shifted FBG-based interferometry system. Sensitivity was compared with different configurations (13). In this context, Wang et

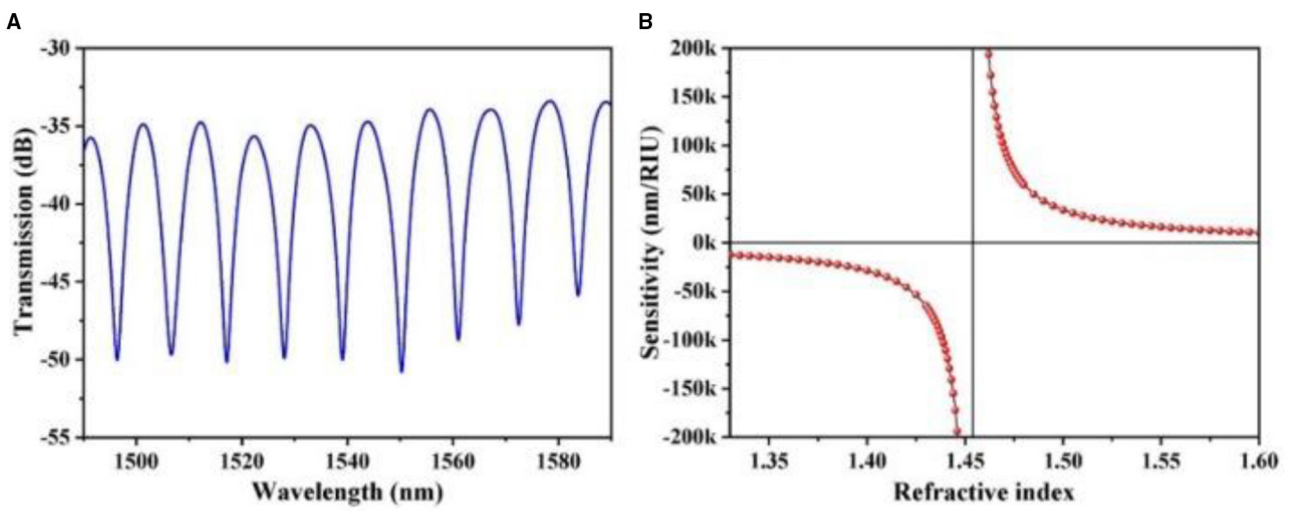


FIGURE 4 | (A,B) Transmission spectra and sensitivity calculation of Machzehnder Interferometer (12).

al. developed an MZI-based sensor for continuous respiratory monitoring. During the development of the MZI-based respiratory monitoring system, it was adequately scrutinised for its respiration rate, tidal volume, and minute ventilation. The designed sensor was constructed with a flexible arch-type structure feature with curvature sensitivity of 8.53 dB/m (14, 15). The study of Zunic et al. designed a MOEMS ultrasonic sensor for photoacoustic imaging. MOEMS sensor was designed and developed with the integration of the MZI. The model was designed and simulated in Comsol Multiphysics. Analysis has been conducted for different materials assigned to layers. Different materials assigned were Si, SiO2, Si3N4, and SMMA. The sensor was designed to operate at 1 MHz (16). The MZI was designed for measuring glucose in serum. The light that was propagating through the sensing system which detected the serum sample with a specific concentration of glucose was analysed. The output power obtained was linear until it reached the value of 1.358 (17).

Optical MEMS Microcantilever

Micro-electromechanical system microcantilevers have wide applications in many biomedical aspects, especially most of MEMS biosensor integrated with microcantilever system because of its flexibility in mass sensing, benefits of EMI insensitivity, and micro-sized fewer circuitry parts (18). Several micromechanical structures were configured with the microcantilever sensor which were widely used for label-free biosensor applications. It has been used in many applications and research studies, including the fabrication of monolithic Wheatstone bridge circuits (18, 19) are the analysis of electrical responses of MEMS piezoresistive microcantilevers. Microcantilevers have been widely used in medical applications such as the induced mass change technique for glucose detection (20–22). Microcantilever sensor work is mainly based on maximum surface stress generation due to applied force at the cantilever tip. Combining optical systems with micro cantilever makes the optical sensor system work effectively for change in deflection and maximum surface stress. In terms of label-free sensing, variety of methods can be integrated with microcantilever sensors. They are (1) photonic crystal structure, (2) MZI, (3) optical ring resonator waveguides, (4) FBG, etc. Label-free sensing uses protocols of change in refractive index due to change in surface density instead of a change in mass. In this review, we discussed microcantilevers, which have undergone micromachining technology. Cantilever deflection is highly sensitive to change in maximum stress developed at the surface of the cantilever. Biomolecules, DNA, and Protein molecules can be captured by the targeted surface of the cantilever coated by the adsorption layer of the material, which increases the surface stress. Adsorption in one side of the cantilever brings out differential stress for each deflection of the cantilever. The electrical and optical methods of cantilever readout methods are in commercialisation. Still, the optical system of the cantilever sensor has more advantages than electrical due to its simplicity, lightweight, small size, EMI Insensitivity, less circuitry complexity, reliability, and linear results (23). In this context, Lee et al. discussed a nanomechanical cantilever sensor integrated with a photonic crystal microcavity resonator. Resonant wavelength shift was observed for each increment of applied force. It has shown a promising future of application in the detection of biomolecules (24). In another work, Xiang and Lee proposed optical readout chemical sensing based on microcantilever sensors. Photonic crystal microcavity resonators integrated with microcantilevers in a fluidic channel for chemical sensing with water and air were investigated. Microcantilever sensitivity in both air and water for the designed cantilever was analysed, wherein minimum detectability obtained in water was less than in the air (25). Microcantilever was used for the detection of specific biomolecules. Overall geometry was designed, allowing simple microfluidic delivery channels to make adsorb to the surface of microcantilever (26). An optical readout mechanism was used for analysing the optical MEMS sensor. Ptolemy-II is a tool considered for experimental validation of designed optical MEMS sensors. Microcantilever was integrated with many other photonic technologies for the development of optical MEMS-based sensors. Photonic crystal microcavity with micro cantilever is applicable in a different biomechanical application. Optimising photonic crystal structure integrated with micro cantilever cavity for label-free biosensing application has more scope in improving sensitivity. Variation of ring thickness and ring radius is an essential factor affecting the overall sensitivity of the ring resonator integrated with the microcantilever. From **Figure 7A**, we can see the microcantilever array is embedded with microfluidic channels for analysing the chemical solution (30, 31). **Figures 5A–E** shows different methods of optical system integration with Microcantilever including a microfluidic channel.

In this context, Steven et al. discussed carbon nanotube-based carbon mechanical resonators. It has been used to detect different molecules in the air. The length of the cantilever ranges from 500 μm to 5 mm. The porosity of carbon resonators varied with a density of 102–104 kg/m^{-3}. Significant applications in biomedical disease detection and pathogens in air and gas environments have been analysed (32). SU-8 polymer material based microcantilever designed and analysed for disease detection. From the evolution of SU8 Polymer-based surface stress microcantilever. The selection of material, geometrical design parameters, and challenges faced in the current market are discussed (33). By considering different materials, maximum deflection for each specific material of microcantilever is analysed. Microfluidic channels have been designed and analysed for blood flow and viscosity generation for the same sensor configurations. Pressure-induced on microcantilever due to an increase in blood viscosity is investigated. The performance of different microcantilever sensors for different materials is carried out. They concluded that SU8 polymer material gives the best displacement and strain distribution range along the microcantilever (34).

Photonic Crystal Pressure Sensor

The photonic crystal is an optical nanostructure that affects the photon motion much similar to the motion of electrons affected by the ionic lattices. Photonic crystal structures are susceptible to change in refractive index by mechanical deformation. A minute change in deflection causes the overall refractive index

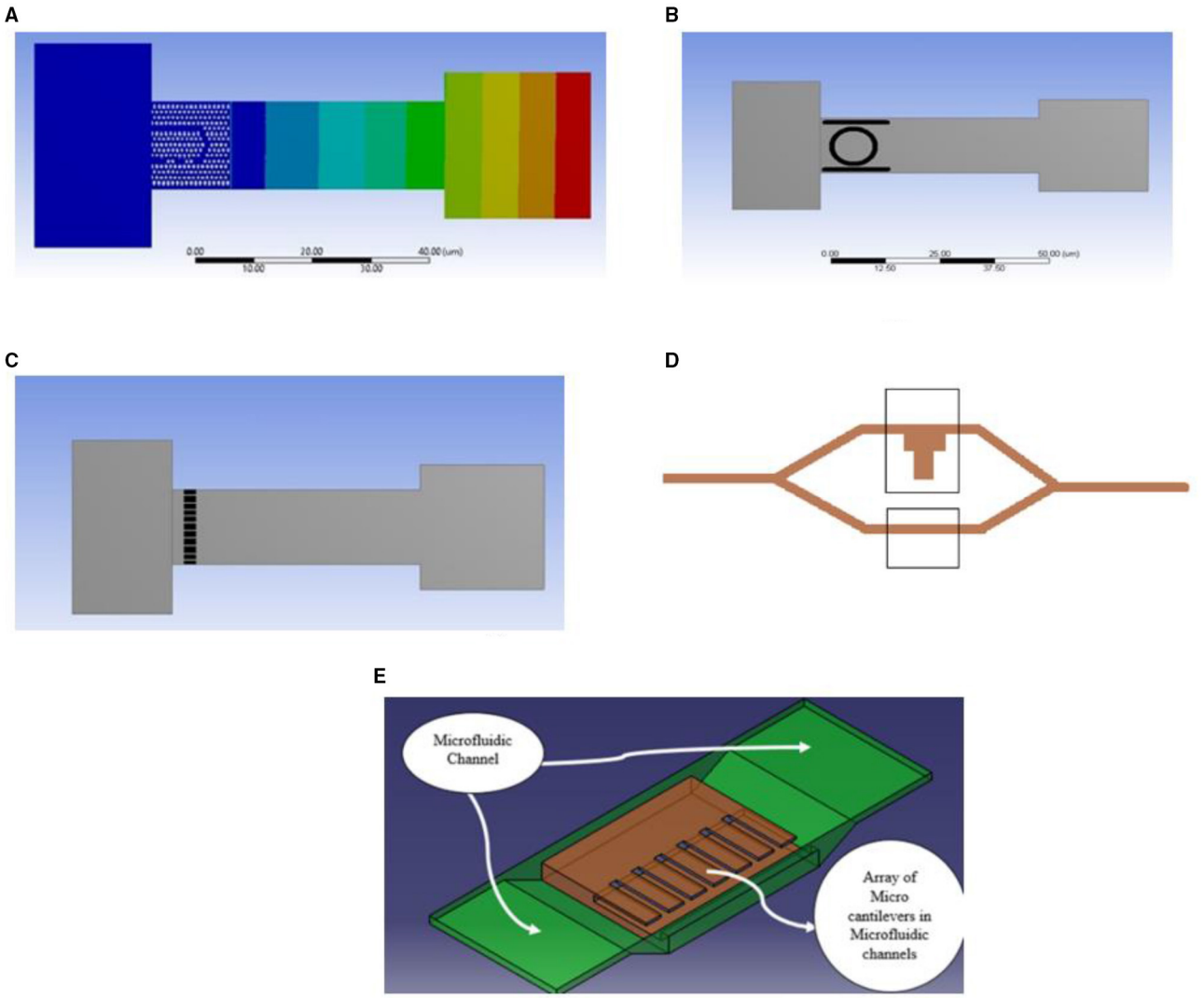


FIGURE 5 | Presents the various Microcantilever: Photonic crystal integrated Microcantilever **(A)**, ring resonator integrated Microcantilever **(B)**, FBG integrated Microcantilever **(C)**, MZI Integrated with Microcantilever **(D)**, array of Microcantilever with the microfluidic channel **(E)**.

of the structure to change. Agarwal et al. discussed a photonic crystal-based micro-opto electro mechanical system sensor in this context. Sensor configuration consists of the integration of a photonic crystal sensing layer in a bridge-type silicon structure. In this context, two cases of sensing structures were investigated, i.e., in the first case, the bridge structure was fixed at one end, while in the second case, it was fixed at both the end. The pressure is applied at the centre of the bridge in the second case. In the first case, the pressure was applied at the free end tip of the sensing structure. The pressure is applied from 0 to 400 MPa for both sensing configurations. The analysis observed that the structure fixed at one end shows higher sensitivity of about 68.296 nm/MPa. Structure fixed at both ends shows a sensitivity of 0.134 nm/MPa. The figure shows the resonant peak obtained for the structure fixed at one end. We can see from **Figure 6** that there is a distinct shift in wavelength for the different pressure range from 0 to 400 Mpa. The wavelength range obtained for structure with one end fixed is 1,550 to 1,535 nm (35). The peak resonant wavelength for each pressure increment is shown in **Figures 6A–C**.

In another work, a hybrid dual-core photonic crystal fibre-based pressure sensor was designed. Elliptical and circular shaped air holes were constructed in photonic crystal fibre pressure sensor configuration. Under different applied pressure, the mode behaviour of the sensing configuration was analysed. The sensor was designed for hydro pressure applications. It was found that a sensor length of 6 cm have sensing range from 0 to 1,000 MPa (37). In this context, Rajasekar et al. discussed a hexagonal photonic crystal ring resonator structure for sensing pressure and temperature. The pressure range considered was 0.04–6 GPa and the temperature range obtained was 5–540°C Bandgap analysis of each sensing configuration was carried out to monitor the range

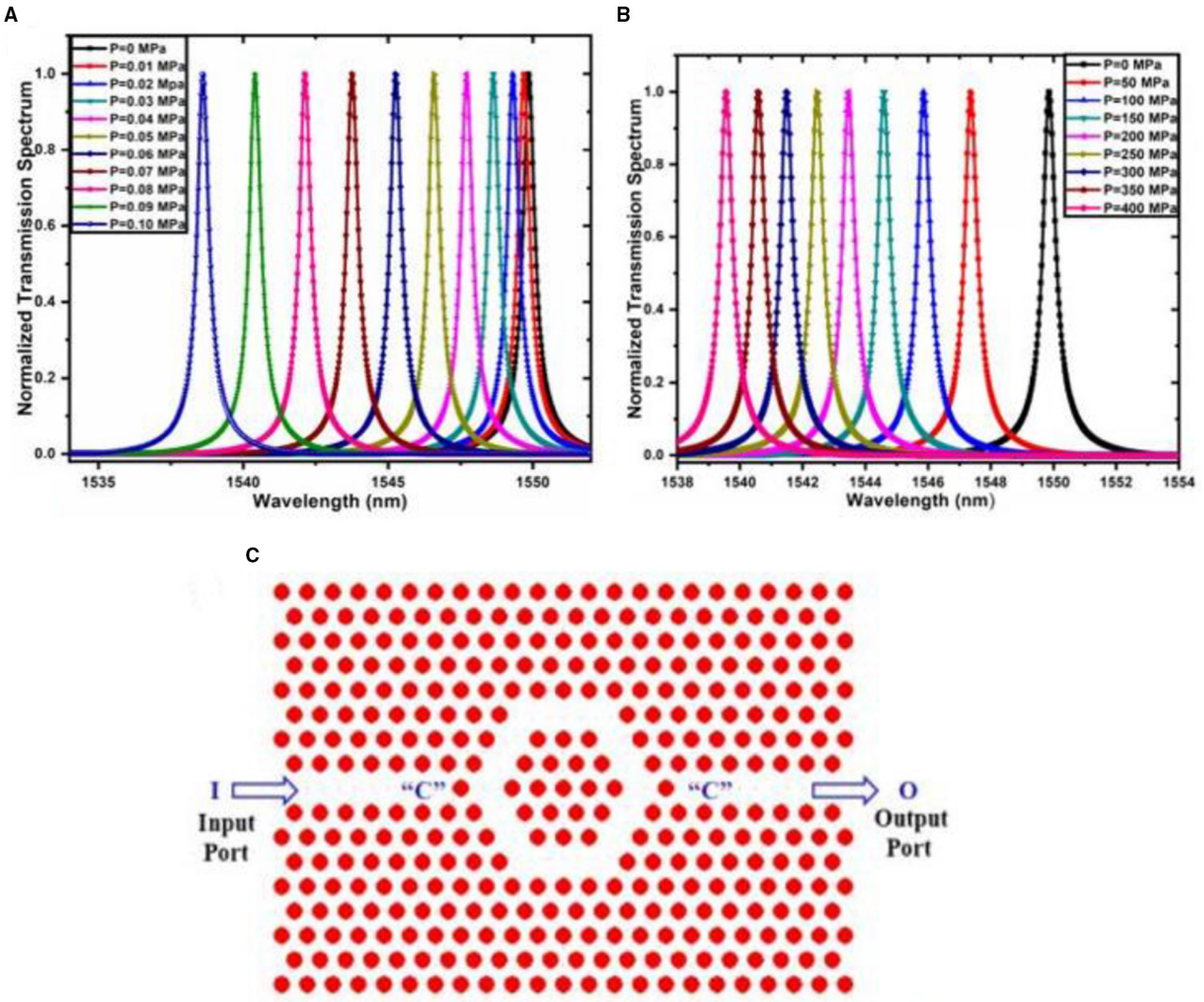


FIGURE 6 | Peak resonant wavelength for sensing structure **(A)** fixed at one end. **(B)** Fixed at both the end (35). **(C)** Hexagonal photonic crystal ring resonator-based pressure-temperature sensor (36).

of frequency. 2D hexagonal photonic crystal resonator arranged with a circular rod in air configuration was considered during the analysis. In another work, a 2D hexagonal lattice of six hexagonal-shaped rings of hexagonal lattice sensing configuration was micromachined on tip of the microcantilever. The concept of using a microcantilever tip is to get a wide range of pressure. The pressure was applied from 1 to 10 Mpa with a resolution of 10 MPa (36, 38). The hexagonal ring resonator has shown better sensitivity up to 250 nm/MPa. The modified refractive index after applying pressure for 2D photonic crystals was calculated by the following Equation 3 (36).

$$n = no - (C_1 + 2C_2)\sigma \quad (3)$$

where no is the refractive index at zero pressure, C1 and C2 are photoelastic constants and σ is the stress generated at the sensing structure due to pressure. **Figure 6B** shows the photonic ring resonator structure.

Fibre Bragg Grating Sensors

Fibre Bragg Grating has a unique characteristic to perform as a sensing system. The core of the FBG is exposed to intense laser light. Refractive index increases as there is an increase in exposure to the laser light. In this context, Werneck et al. discussed the basic principle of FBG sensors. A periodic change in the refractive index in the core segment of optical fibre was grating fibre. Refractive index variation due to applying pressure or temperature led to pitch change in FBG. The most common choice of fibre optic-based sensor is FBG sensors. The ultrasonic laser inscribing process generated the periodic refractive index on a core segment of optical fibre. An FBG suitable for measuring strain in a product or device has experienced pressures

throughout the process. The multiple FBG sensors could be embedded under any structure to get the spectrum shift from the FBG interrogator. The most application of FBG sensors was based on the structural health monitoring system. In structural health monitoring, FBG sensor placement is the most crucial part of the research. Since there is an application of the thermal or structural load on the FBG sensor, installing an FBG sensor inside the structure requires prior knowledge of the structural behaviour. The FBG sensor placement study was conducted in many mechanical applications like the placement of fibre in the tensile test specimen, placement of fibre in wings of the aeroplane for structural health monitoring, or in the case of rail monitoring, during the movement of train wheel for a specific period. From a biomedical point of view, most blood pressure measurements and temperature measurements were in progress with the FBG system. In many applications, the FBG sensor was used as non-invasive. Most FBG Interrogators will provide the raw data in the form of voltage; the same is converted in wavelength for each change in microstrain and feed rate of FBG sensor (27).

The study of Bonefacino et al. discussed ultra-fast polymer optical FBG inscription for medical devices in this context. The fibre consisted of polymethyl methacrylate material. The extraordinary result was obtained during the experimentation. New dopant diphenyl disulfide causing positive refractive index change was utilised during the process. This made sensitivity high when FBG was considered with PMMA to compare with silica fibre. The sensitivity of the FBG sensor obtained was -55 pm$^{\circ}C^{-1}$ (39). The strain from the FBG sensor was also used to calculate the curvature and torsion of fibres. The catheter was used for reconstruction techniques. The FBG embedded catheter was used as a shape sensor. The maximum reconstruction error found was 1.05 mm. The authors concluded that shape sensing with flexible biomedical instruments is feasible with the FBG sensor (40).

Along with these applications, the uses of FBG in civil engineering applications, biomedical applications, military, and marine applications were discussed. Depending on the type of application, FBG is designed with long or chirped grating, which is uniform and tilted (40, 41). In this context, Zheng et al. discussed 3D hydrogel actuators driven by light and generated by two-photon polymerization. The photoresist of gel was produced by Fe_3O_4. The size of the hydrogel actuator was 26 μm and the response time of the sensor was 0.033 s. The designed actuator has good controllability and repeatability (42). On-chip optomechanical sensing configuration has been established. Sensor configuration consisted of a dual-channel integrated waveguide and directional coupler. The dynamic range of the sensor was 30 nm with displacement imprecision of 45 fm/Hz (43). In another work, Linh et al. proposed an FBG-based optical pressure sensor. The pressure sensing element was constructed with Fabry-Perot based pressure sensor. Fabry-Perot pressure sensor with a different cavity length was optimised. Obtained pressure sensitivity was -0.672 rad/MPa. The proposed sensor was having a very high-temperature sensing ability (44). In another work, optical fibre-based pressure sensing application system was developed for pulse transit time measurement. Pulse transit time measurement system is helpful in cuffless blood pressure measurement. Electrocardiograms and photoplethysmographs had been monitored with a pulse transit time system. Plastic optical fibre was deployed as a transport medium in the chest part of the human to monitor the pulse pressure. Research work carried out here considered pulse transit time (PTT) in three different positions, i.e., finer, wrist, and foot (45). Complementary metal oxide semiconductor (CMOS) MEMS-based membrane nanomechanical sensor is designed and developed for molecule detection. Results shows that small molecules significantly impact the human body, wherein detecting such a molecule is challenging in many sensing systems. This membrane bridge sensor is used to detect phenytoin in different concentrations. The detection limit of this sensor is 4.06 $\pm$ 0.15 μg/ml. It is used as a biosensor as well chemical sensor. In this context, Raghuwanshi et al. carried out a sensitivity analysis for TiO_2-coated FBG sensor for chemical detection. A TiO_2-based FBG sensor was developed with Cu-Vapour-based laser writing. The thickness of the coating has been optimised to get better sensitivity and FBG design parameters. The sensor was able to detect the minute change in adulteration with accuracy 0.01 ppm (46, 47). In the study of LV, C and co-authors, they discussed invasive FBG based pressure sensors for palpation of tissue surgery. The finite element method has been used to analyze the static and dynamic performance of the sensor configuration. Optimum sensor design output has been given to 2.55 mN. The pressure range considered was 0–5 N (48). In another work, the effectiveness of sensing configuration has been presented by Khan et al. for the pose measurement system of different medical instruments by FBG sensor. The results of the FBG sensor were converted to strain. Strain calculations were made by comparing strain gauge parameters. Mean error in the orientation and position was observed to be <4.69 mm and 6.84°. Different poses of catheters and many other medical instruments by experimental data were obtained (49). In this context, Antunes discussed the mechanical testing of dental resin. In real-time usage of dental resin, resin materials had undergone failure due to the shrinkage polymerization. This caused the breakage of gaps, which interfaced the dental resin with the tooth. A laboratory experiment was conducted to cheque the ability of different resin materials and their failure point by using an FBG sensor. The FBG sensor was placed in the dental model (28). Variation of strain concerning time with FBG sensor is shown in **Figure 7B**. **Figure 7A** shows the FBG sensor processing system. Performance comparison of different sensors is shown in **Table 1**.

Figure 7C shows the temperature variation during resin polymerization in dental material. The shift in wavelength was obtained for change in these parameters. Strain dependence of FBG sensor can explain with the following mathematical expression (4) and (5) (29).

$$\Delta\lambda/\lambda_0 = n_{eff}/\Lambda_(eff_\Lambda) = (1 + (1/n_{eff})\delta n_{eff}\delta\varepsilon)\Delta\varepsilon = (1 + p_e)\Delta\varepsilon \quad (4)$$

$$\Delta\lambda\lambda_0 = k\Delta\varepsilon \quad (5)$$

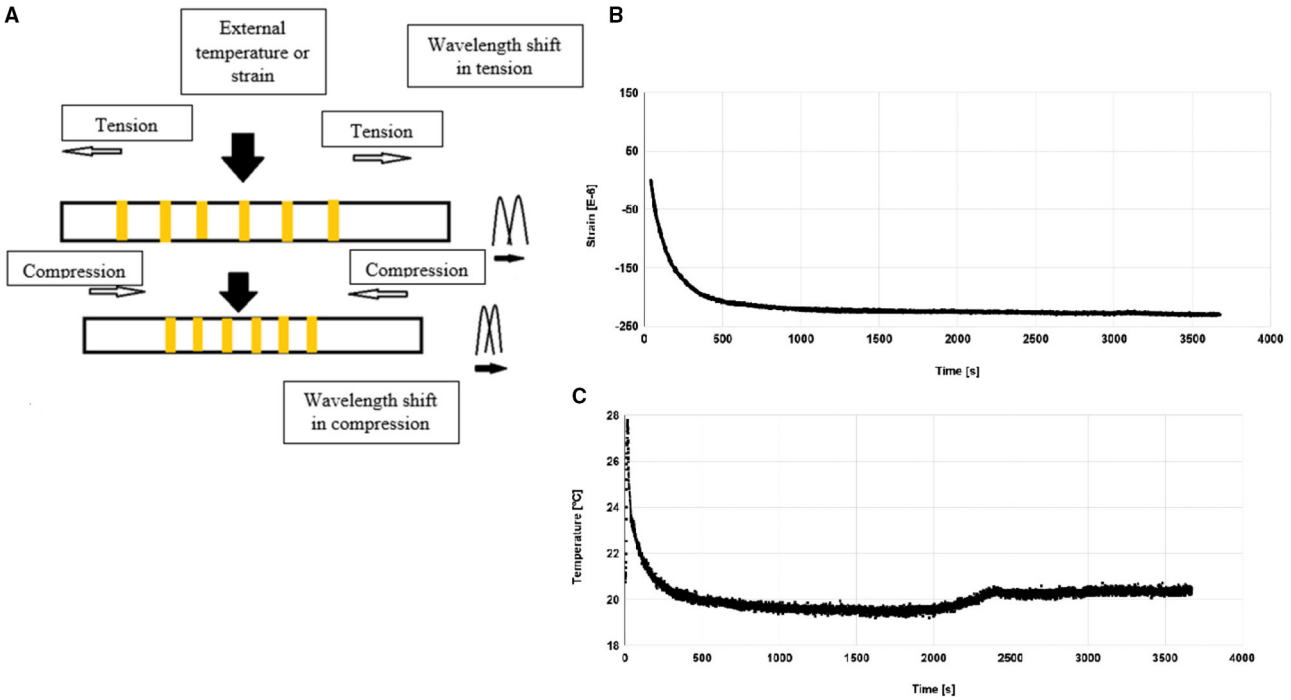


FIGURE 7 | (A) Fiber bragg grating working process during the application of pressure (27). **(B)** Strain evolution of bulk flow resin during polymerization (28). **(C)** Temperature evolution of bulk flow resin during polymerization (29).

where k = k factor of bragg grating, pe = photo elastic constant = 0.21, and $\Delta\lambda/\Delta\varepsilon = 1.2$ pm/(μm/m) - Strain sensitivity for FBG at 1,550 nm

Fibre Bragg Grating as temperature dependent has below expressions;

$$\begin{aligned}&\Delta\lambda/\lambda_0\\&= \Delta n_{eff}/\Lambda_(eff_\Lambda) = (1/\Lambda(\Lambda/\delta T) + (1/n_{eff})\delta n_{eff}\delta T)\Delta\varepsilon\\&= (\alpha+\zeta)\Delta T \qquad (6)\end{aligned}$$

α – coefficient of thermal expansion
ζ- thermo-optic coefficient
$\alpha = 0.55 \times 10^{-6}/°C$
$\zeta = 5.77 \times 10^{-6}/°C$.

Application of Optical MEMS

Miniaturisation of devices or various micro mechanical elements is a critical concept in developing new optical MEMS technology. Optical MEMS are the microdevices that measure and manipulate various mechanical parameters such as pressure, temperature, shape, stress, strain, etc. Optical MEMS has plenty of applications in present days, e.g., in industry, biomedicine, automobile, and aviation. In this paper, we selected the critical application of optical MEMS in the biomedical field. The primary application discussed here was dental, human spinal disc, optical MEMS in the human respiratory system, the optical fibre in gait motion, blood pressure measurement, and bladder pressure monitoring, among others. Detailed discussion on the advancement of optical MEMS in the stated different applications is presented below. Most of the applications had used optical fibre-based FBG sensing systems (Section Fibre Bragg Grating Sensors) for real-time applications. Microcantilever integrated different types of photonic crystal structures (Section Optical MEMS Microcantilever) were modelled and simulated for various applications. **Figures 8A–D** shows the Optical MEMS sensors applications in the human body.

Dental Application

Incorporating the optical sensor for the dental caries monitoring in the dental application is a new innovative technique. Evaluating the dental filler materials and monitoring the structural behaviour of braces and aligners are essential criteria in dental health monitoring. Recent dental structural behaviour evaluations monitored by the FBG technique were discussed in this review. Evidence suggests that orthodontic tooth movement and biting force can be monitored with photonic MEMS technologies. The work of Silvia et al. proposed a 3D intraoral scanner to map the 3D image in dental restoration accurately. Dental caries is a major problem in dentistry. Global disease study in the year 2016 has been predicted that in the total world of population, wherein 3.5 billion people have been remarked with oral diseases such as tooth decay. Cracked tooth caries are complex tissue lesions that severely affect the quality of life of a person. It has remained a mystery to solve the non-invasive method to detect the cause of caries and crack of the tooth in humans.

TABLE 1 | Performance comparison of selected sensors.

Type	Configuration	Strategy	Mode	Detection limit	Sensitivity	References
Interferometer	MZI	Three core fibre for measurement of directional bending	TE	n.r.[a]	Bending sensitivity −15.35 nm/m^{-1} and 3.11 nm/m^{-1} Temperature sensitivitiy 0.043 nm/°C and 0.041	(3)
	Diaphragm Integrated MZI	Analysing to light intensity with diaphragm deflection	TE	n.r.[a]	−0.2 pm/Pa	(5)
	MZI	Lab on chip biosensor with MZI Phase modulator		<0.05 L/min), a high resolution (0.005 L/min), and a fast response time (12 millis	n.r.[a]	(6)
	MZI	Methylcellulose coated MZI for humidity sensing	TE	45–85% RH	0.094 dB/%RH	(10)
	MZI	MZI Ultrasound Sensor	TE	0.38 $mPa/Hz^{1/2}$	0.62 mV/Pa	(11)
	MZI	MZI temperature sensor	TE	n.r.[a]	21.2 nm/C	(12)
	MZI	MZI Respiratory monitoring	TE	n.r.[a]	Sensitivity upto 8.53 dB/m-1	(14)
Photonic MEMS /Microcantilever	Photonic Microcantilever	Photonic Microcantilever for chemical analysis	TE	0.6 mum and 0.0098% in water and 0.812 mum and 0.0144% in air,	n.r.[a]	(25)
	Optical MEMS microcantilever	Various shape microcantilever integration with Optical MEMS	NA	81.55 μg−28.39 mg (polyimide material –Triangular cantilever. 50.97 μg−23.996 mg (Parylene material	n.r.[a]	(30)
	MOEMS	MOEMS nanomechanical sensor	TE	0–0.10 MPa;(One end fixed) 0–400 MPa; (both end fixed)	68.296 nm/MPa (fixed one end) 0.134 nm/MPa (Both end fixed)	(35)
	Photonic crystal fibre	Hybrid dual-core photonic crystal fibre	TE	0 to 1,000 Mpa	−11.6 pm/Mpa.	(37)
	Photonic crystal	Pressure and Temperature sensor with hexagonal resonator	TE	0.04 to 6 GPa	n.r.[a]	(36)
	Microcantilever Hexagonal ring resonator	hexagonal ring and micromachined cantilever tips on 2D silicon photonic crystal	TE	1–10 MPa, 10–100 MPa, 1-MPa −20MPa, 20 MPa−10 GPa	n.r.[a]	(38)
Fibre Bragg Grating	FBG sensor for biomedical	Polymer optical fibre Bragg grating analysis	TE	n.r.[a]	54.2 pm %RH−1 ± 0.14%RH	(39)
	FBG for biomedical	Integrated nano-optomechanical displacement sensor		Displacement imprecision of only 45 FM/Hz1/2 as well as a large dynamic range (>30 nm).	n.r.[a]	(43)
	Fibre pressure sensor	Multi-Point Optical Fibre	TE	n.r.[a]	−0.672 rad/MPa	(44)
	CMOS based Nanomechanical sensor	CMOS MEMS-based nanomechanical sensor for molecule detection	NA	4.06 ± 0.15 μg/mL	n.r.[a]	(45)
	FBG force sensor	FBG force sensor for tissue palpation	TE	0–5 N	n.r.[a]	(48)

[a]*Not reported.*

Nevertheless, early detection of caries is made possible by FBG based optical sensor (53). **Figure 7** shows the optical MEMS application overview.

In another work, Pyo et al. discussed the feasibility of using optical coherence tomography in dentistry. Analysis has been carried out by changing the fringe numbers, clarity of dental images has been investigated with optical coherence tomography (OCT). OCT has more advantages for exploring a more complex tooth portion (50). Hybrid technology called photoacoustic imaging high-resolution ultrasound could also project deep tissue with high contrast optical images. **Figures 8A,C**, show dental scanned images. Strain patterns for different orthodontic devices were evaluated. Regarding this, an orthodontic appliance with a disjunctor bar on the left side and without a disjunctor bar

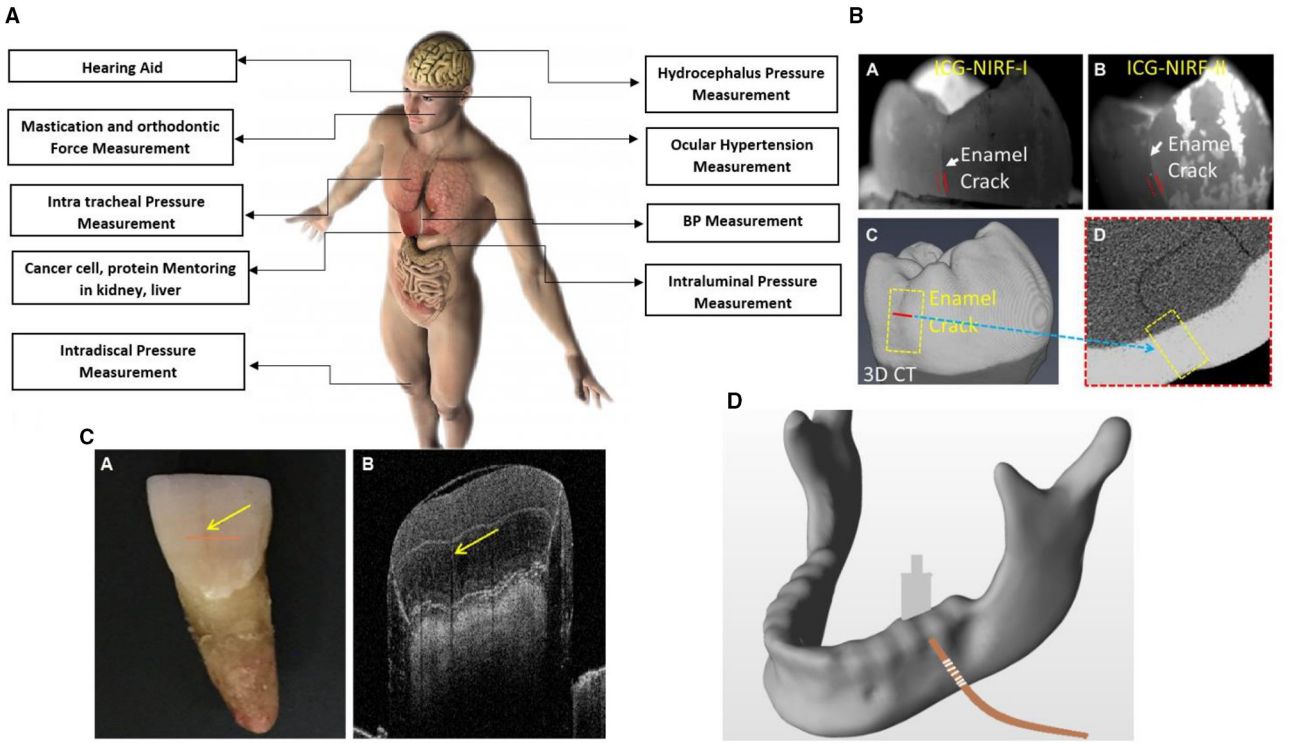


FIGURE 8 | (A) Optical MEMS sensors application in the human body (50). **(B)** Dental X-Ray imaging of cracks by green near infrared imaging (51). **(C)** Dental OCT imaging (53). **(D)** FBG integration with human mandibular (52).

was compared for strain pattern obtained for transferred forces to the dental arch, teeth, or maxilla (54). **Figure 8D** shows the placement of the FBG sensor human mandibular model. In this context, the FBG sensor was used to measure material expansion because of a dental implant. The effect of thermal strain on FBG sensors has been widely studied. The hardening phase of the material expansion process has been investigated concerning the optical MEMS sensor. Electromagnetic interference (EMI)-insensitive, small dimension bite force device has been developed by FBG technology for measuring. Obtained results showed better resolution compared with the traditional devices available in the market. The optical system was also used in strain measurement of the tibia in bone for decalcification. The different bone sample has experimented for different percentage of decalcification. Since there is a loss in calcium percentage in bone, an increase in strain behaviour was observed. Better resolution over the result was obtained compared with QCT, QUS, and DEXA. These traditional techniques had the drawback of inaccurate values and exposure to radiation (51, 55). In another work, the surface roughness of rot carriers was analysed using optical profilometry. A total of 20 extracted teeth from a patient was examined for root caries, and surface roughness was analysed with profilometry. This clinical study was carried out to protect the non-cavitated surface from cavitation by prior accession of the status of a tooth. In the dental clinical examination, there was no clinical test process to detect hard tissue. This need in the dental field is the reason for the development of optical polarisation imaging systems. This imaging system could assist in dental calculus caries and cracked tooth (52, 56).

Strain Monitoring Human Spinal Disc

The degenerated human intervertebral discs exhibit more compressive stress, tensile stress, and strains. Load distribution on the spinal disc is important for the proper functioning of the lumbar motion of the human body. In this context, Newwell et al. evaluated the use of the FBG sensor in human spinal cord mechanics. Understanding the pressure distribution of the intervertebral disc is a major key point in the clinical diagnostic of disc strain. The biomechanical structure of the spine consists of the intervertebral disc, as shown in **Figures 9A–C**, between superior and posterior parts of the spine consist of an intervertebral disc. In the study of Newwell et al., biomechanical action took place between these parts. Finite element mandrel was developed with encapsulation of FBG sensor within it. FBG sensor was incorporated with the intervertebral disc part. Compressive or tensile force applied on the intervertebral disc has generated the strain. Encapsulated FBG sensors had undergone pressure that brought the change in the grating period. Functional units of the spinal cord experienced compressive load. FBG sensor

[1]Istock. Avaialble online at: https://www.istockphoto.com/in (accessed July 20, 2019).

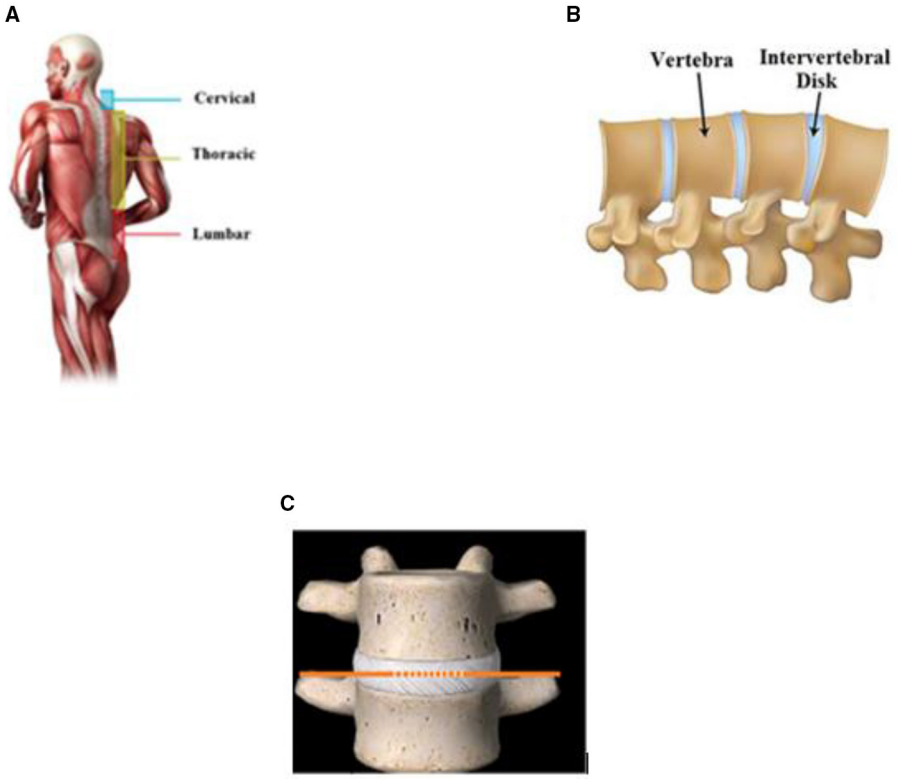


FIGURE 9 | Spinal cord parts of human with integrated FBG sensor **(A)** (57). Intervertebral disk in the spinal cord **(B)** (52). FBG sensor integrated with intervertebral disk **(C)** (57).

incorporated in intervertebral disc and linearity, sensitivity, and hysteresis for each compressive stress was investigated. The porcine intervertebral disc (IVD) pressure has been validated with an FBG sensor for accuracy. Needle mounted strain gauge sensor was inserted into the nucleus. Even though the invasive measurement method was divisive, this method could undergo accurate pressure measurement in the IVD. Functional units of the spine were compressed for 0–500 N force. The substantial potential has been exhibited by FBG sensor with IVD pressure monitoring by sidestepping the controversial nature of invasive methods (58, 59). In another work, Yamaji et al. evaluated the device for analysing the motion of the lumbar. The device consisted of stretchable strain sensors. It was observed from the clinical survey that lower back pain was common in most people in the world. In this study, the stretchable device for lower back motion was analysed. Six strain sensors were used in the human spinal cord in the non-invasive method. Sitting postures were the major cause of most spinal cord diseases. In such cases, detection of the wrong posture could give major advantages. Many wearable devices had been investigated for different applications, wherein the sensitivity of the sensor achieved as high as 0.2 nm/$\mu\varepsilon^{-1}$ (60, 61).

[2]Available online at: https://grabcad.com/ (accessed July 27, 2020).

Monitoring Angular Motion of Joint and Gait Analysis Application

Diagnosing issues causing pain during walking or implementing or evaluating pain for treatments is paramount in physiotherapy treatments. Monitoring the angular motion of the joint helps in restoring correct patterns and abnormalities. In this context, Padma et al. proposed a goniometer device with FBG encapsulated for dynamic angular motion of joint. Rotational motion of join in the affected human body was converted into strain variation. Developed FBG sensor was able to detect the range of angular motion from 0 to 200°. The resolution obtained was in the range of 0.06°. Elbow joint and ankle joint measurements were carried during the experimentation. The sensor device was used to measure the angle between the shin and thigh for various cantilever strain variations. Knee angle measurement device was also validated for a range of motion of joint angle in gait analysis with traditional sensors. Sensitivity switching mode was made available to obtain various levels of sensitivity action mode. Low sensitivity mode, medium sensitivity mode, and high sensitivity mode were developed in the device. They also used polymer optical fibre for joint angle measurement. The obtained results were compared with a resistive goniometer and polymer optical fibre sensor (57, 62). In another work, an electro goniometer was used to measure the joint angle. The goniometer was embedded with an FBG sensor to get the angular motion details. However, the drawback

of this device was the fixed hinge and fixed centre of rotation which caused a problem during the measurement of angle in joint with more bend. Alternative fibre optic-based device with the technique of intensity modulation and laser beam methodology was developed. The prototype developed had a range of angular rotation up to 90° with a standard deviation of 0.1°. The developed sensor has high reliability, sensitivity, speed, and flexibility (63). In this context, JR Galvao discussed gait analysis with FBG sensor. Gait analysis is the study of the motion of a human or animal body using the eyes of the observer through various study devices for the measurement of motion with body mechanics. Sport and medicine are the widely used area of gait analysis. The walking process, steps in walking, body posture, and walking velocity are the different parameters depending on the gait analysis result. Different gaits of horses had been studied using the instrumentation of the FBG sensor. Early injury diagnosis of horses and retirement of horses were investigated in the said study. Initially, the FBG sensor was used without any encapsulation and embedded in each limb of the horse. In the second step, FBG was encapsulated with composite fibre and embedded in the limbs. The second step involved digital image processing techniques to get proper images of compressive or tensile forces. Both FBG sensing and digital image processing had shown good agreement for biomechanical study and evaluation of FBG sensors for horse dynamic training and competitions (64, 65). Furthermore, the study of Domingues et al. discussed an insole FBG-based optical fibre pressure sensor. Internet of Things technology was incorporated with this application. An optical fibre-based foot pressure system was integrated with the Internet of things for e-health assessment. Accuracy of the system was varied with normal walking and patients with diabetes. Polymer optical fibre was embedded in the foot plantar surface. An array of polymer optical fibre has placed foot insole and experimented for different gait cycles. Four different FBG sensors characterised by four different wavelengths had been used. The FBG sensor was positioned at different positions of insole, i.e., first, second, third, fourth, and fifth metatarsus. Finite element analysis (FEA) was carried out for the insole model. The high sensitivity of 0.000492 nm/KPa was obtained for the developed system. The insole was embedded with multiplexed FBG-based sensor. The designed system was allowing complete gait motion assessment with a multiplexed FBG sensor network. Assessment results of the insole would be sent to a physiotherapist to monitor the conditions of the patient (66–69).

Respiratory Pressure Monitoring

Monitoring of respiratory system reduces the severe illness during consideration of clinical response. Exploring the possibilities of good respiratory monitoring sensors such as the FBG system is vital in the present scenarios of chronic respiratory diseases. In this context, Padma et al. discussed a breath pattern analyzer based on the FBG technology for communication assistance for people having restricted communication ability. In emergency medicine for people with spinal cord or brain injuries, or patients in the intensive care unit, critical life-supporting system is required. The FBG breath pattern analyzer was used for the patient to communicate with the outside world, where specific breath patterns were trained by FBG breath pattern analyzer. During emergency assistance, the patient could be communicated for specific purposes. The developed device was an alternative to the brain-computer interface, where specific instruments were used to obtain the brain responses (70). In this context, Krehel developed a textile-based respiratory system. In this system, highly flexible polymeric textile fibre was embedded with an optical fibre pressure sensor, and the pressure applied on optical fibre has shown a distinct shift in wavelength. The device was analysed with different setups with an optical fibre pressure sensor. Half oval with a length of 60, 180, and 240 mm, as well as half oval with additional bends and with additional bend and cross fibre is cosnidered in the design. The pressure was monitored with a human torso. This comparison with different setups was made to assess the utility of this monitoring device (71). In another work, Nishiyama et al. discussed and evaluated respiration and body motion during sleep with FBG technology. This investigation is essential to understand the human breathing pattern during sleep and detect the motion of the body, sleep habits, etc. The optical fibre pressure measurement system was embedded in a conventional bed. Pressure changing due to breath and roll over in bed was monitored throughout the process using plastic optical fibre-based respiratory monitoring system. Plastic optical fibre was used as a sensor to measure the rate of breathing. This optical fibre pressure sensor has been placed in tape form in the chest of the patient to monitor respiratory movements. The input and output signal from the optical fibre pressure sensor was obtained as intensity modulation. The pressure sensor was tested in a chest in normal status, sitting, and lying condition. The proposed sensor has provided different possibilities in respiratory monitoring within the MRI system (72, 73). In this context, Massaroni et al. discussed the method of measuring respiratory rate by the contact-based method. Different respiratory rates were recorded with the help of sensing element. The complete details discussed were about the contact-based methods, monitoring airflow pressure, air humidity, air movement, chest wall movement, and sound. Differential flow metre, turbine flow metre, anemometer, and fibre optic sensor were used during the process. MEMS-based pressure sensor for simultaneous monitoring of breath pressure and pulse wave was analysed. The real-time experiment was conducted on nasal pressure to monitor the pressure developed at the nasal airways. The MEMS-based pressure sensor pad was attached to the human nose. Eyeglasses was attached the sensor pad to the nasal part (74, 75). In another work, polymer optical fibre-based respiratory rate monitoring sensor has been analysed. In this work, a stretchable strain sensor was made wearable around the stomach. During respiration, inhale and exhale process, the stomach would expand. The sensor would measure the strain generated by expansion. Infrared light was passed through the stretchable grating fibre (76).

Photonic MEMS Microfluidic System

Optical MEMS with microfluidic concepts involves techniques of lab on chip-based sensors. The system consists of optics and fluidic systems with mechanical actuation. Microfluidic

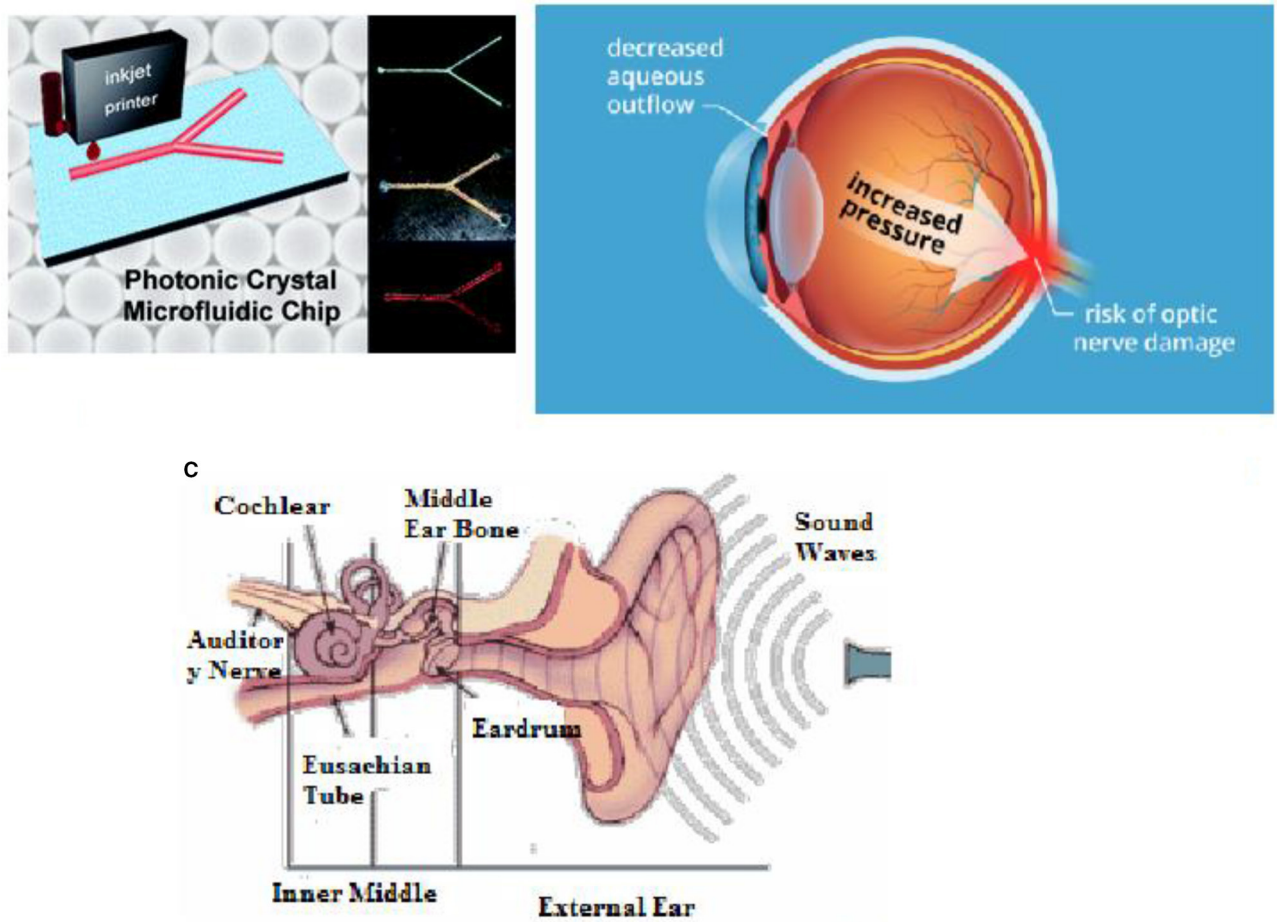


FIGURE 10 | (A) Microfluidic channel with photonic crystal integration (81). **(B)** Intraocular pressure development in the eye (82). **(C)** Human cochlear system (83).

channels are a vital component of photonic MEMS for any biosensing applications to introduce any type of fluid for sensing. In this context, Jeiang et al. discussed an FBG-based Fabry-Perot pressure sensor in a lab on a chip that included microfluidic channels. As the different polystyrene microparticle size varies, the wavelength shift was observed. Sensor systems presented ease of fabrication methods and label-free detection methods. The sensor was demonstrated in polymethyl siloxane chips. It was also monitored that experiments and analytical results fitted very well within the context and avoided the waste of samples (77). In another work, the FBG sensor was integrated into the microfluidic channel. It provided opportunities for dynamic measurement of the physical and chemical properties of fluid passing through the channel. In another work, Weizhi et al. discussed a photonic crystal-based microfluidic channel fabricated with the inkjet printing process. Photonic crystal with the different coloured colloidal microfluidic channel was successfully integrated into a single chip. The fabricated microfluidic channel has the advantage of rapid fabrication possibility, and colourimetric was the add-on properties of photonic crystal based microfluidic channel (78, 79). In this context, Nunes et al. discussed silica-based 1D photonic crystal in opting fluidic systems. It comprised of photonic crystal with the integration of a microfluidic channel. By considering the nanopillar configuration of the photonic crystal, biofluid was injected into the microfluidic channel during the analysis. In the stage of the simulation, the refractive index was taken into account. The shift in resonant wavelength has been observed during the propagation of light due to a change in the overall refractive index profile of the sensing structure with the microfluidic channel. The author has proposed the sensor for biochemical applications (80). **Figure 10A** shows the integration of microfluidic channels with photonics. In another work, Khoshnoud et al. discussed the use of MEMS-based biomedical sensors.

Micro-electromechanical system devices are used to detect triglycerides, glucose, and are used for the measurement of pressure during surgeries, counting the blood cells, sensing in the thermal fluid, and electromagnetic fields (84, 85). In

TABLE 2 | Key applications of Optical MEMS in biomedical.

Medical area	Type of modulation	Place	Monitoring strategy	Technical feature of sensor	References
Dental	Optical Coherence Tomography	*In vitro*	Carries, cracked tooth	B-mode PAT is good at providing microstructural properties. S – Mode PAT provides a proper imaging system of the crack lesion.	(53)
	Fibre Bragg Grating	*In vitro*	Force monitoring in maxilla model	Force range detected 0–20 N	(51)
Intervertebral disc	Fibre Bragg Grating	*In vitro*	Pressure monitoring in intervertebral disc	Pressure range between 0 and 500 N	(59)
Lumbar motion	Fibre Bragg grating	*In vitro*	Lumbar stretch monitoring	Rotation angle upto 0–35°C	(60)
Low back motion	Optical fibre	*In vitro*	Low back movements monitoring	Sensitivity 0.20 nm·mε^{-1}	(61)
Join angle	Fibre Bragg grating	*In vitro*	Goniometer	Joint angles in the range of 0–200° with a resolution of 0.06°	(65)
Insole pressure	Optical fibre	*In vitro*	Insole force monitoring	Sensitivity 11.06 pm/N	(66)
Communication assistance	Optical fibre	*In vitro*	Using breath pattern analyser	Average accuracy of the device obtained 90%	(67)
Pulse wave and respiratory	MEMS + Optical	*In vitro*	Pulse pressure monitoring	Minimum detectable pressure 0.01 Pa	(74)
Microfluidic	Photonic crystal	*In vitro*	Fluid flow sensing	Sensitivity 836 nm/RIU	(87)
Intraocular	Optical MEMS	*In vitro*	Glaucoma Monitoring	accuracy of 0.29 mm	(90)
Oximetry	Optical Fibre	*In vitro*	Oxygen monitoring by the contact force	Detection limit of 5-15KPa	(96)
Pulse wave monitoring	Optical fibre and FBG	*In vitro*	Pulse wave signals	POF-FBG APG correlation is distributed from 0.54 to 0.72	(82)
Urodynamic measurement	Optical fibre pressure sensor	*In vitro*	Urodynamic Pressure	0.1 cm H_2O (∼10 Pa), a stability better than 1 cm H_2O/h	(98)

this context, Quack and Sattar discussed a nanoscale photonic integrated circuit to explore optomechanics. In this work, the authors mainly focused on reducing power consumption due to the movement of mechanical parts within the system. Due to the minor modulation in the effective refractive index, it was possible to get the required consumption of power. The authors also discussed the recent achievement of MEMS enabled silicon photonic systems (86). In another work, an optofluidic pressure sensor integrated with a microfluidic channel was developed. Optofluidic pressure sensor made up of Fabry-Perot resonator system was integrated with a fluid flow wall. The sensor obtained high-quality factors and high sensitivity with a good detection limit. The sensitivity of the sensor was 12.46 nm/MPa with a detection limit of 8.2 bar. A silicon photonic mid-infrared system with microfluidic channels was integrated. In this system, silicon on the insulator layer was developed. The author focused on increasing sensitivity by increasing the length of the interaction of light with structure, intensity, and reducing noise (87, 88).

Measuring Intraocular Pressure

Intraocular pressure (IOP) monitoring is an essential aspect of the human eye to avoid the risk of glaucoma and to provide clear vision to a person. Pressure build-up within the eyes optic nerves is the leading cause of glaucoma. Glaucoma damages the optic nerve, and if untreated, may lead to vision loss, which then makes a person blind. In this context, Lee et al. developed a micro-sized optical implant for *in vivo* intraocular pressure monitoring. Sensors were mounted in intraocular lenses for continuous monitoring of pressure. Sensors were tested with New Zealand white rabbits, and in this way, continuous intraocular pressure measurement was made possible. In another work, intraocular pressure sensor was used with the technology of a flexible photonic crystal membrane.

Periodically, micro-structured 1D photonic crystal on polydimethylsiloxane was considered in the experimentation process. The estimated limit of detection was about 160 Pa. The micro-size compatibility of these sensors has been rendered as a favourable application for glaucoma. **Figure 10B** shows intraocular pressure development. IOP sensor was integrated into keratoprostheses, and the proposed sensor has given 2 mm/Hg accuracy (89–92). In the study of Han et al., they discussed an optical aberrations effect on intraocular pressure measurement with small optical elements in this context. Elevated intraocular pressure is a significant risk factor for glaucoma. To monitor this, the implantable optical pressure sensor was incorporated *in vivo* with rabbit eyes. Optical aberrations relationship with intraocular pressure sensor performance was monitored. It was concluded that sensor readout accuracy decreases as the magnitude of optical aberrations increases. The accuracy of signal to noise ratio obtained was 0.58 ± 0.32 mmHg and 15.57 ± 4.85 dB[3]. In another work, Truong et al. developed a portable handheld reader for monitoring intraocular pressure. Reflected light

[3]Available online at: https://www.allaboutvision.com/conditions/hypertension.htm (accessed July 27, 2020).

from the intraocular pressure sensor forms fringed and it was captured using a camera. Three different handheld readers were used to analyze the better quality fringes and patterns. It was concluded from the results that DSLR based reader has high-quality fringes compared with others. Monitoring of intraocular pressure with optical coherence tomography is a novel technique. Hypertension can lead to glaucoma. Frequent control of IOP monitoring corneal deflection either by direct contact or non-contact process is the current procedure involved in clinical treatment. The relationship between OCT images and the IOP sensor was developed by capturing the bovine eye (81, 93).

Blood Pressure Measurement

Daily monitoring pulse rate is essential to success to meet fit and healthy lifestyles, especially for people with cardiovascular diseases or weight loss. FBG based system is widely applicable for monitoring pulse signals because of its micro size and high sensitivity to external pressure. The study of Katsuragawa et al. evaluated FBG based non-invasive blood pressure sensor. Systolic pressure and diastolic pressure pulse were measured in a single FBG pressure sensor. FBG sensor was kept at the right wrist and right elbow. Results were validated regarding blood pressure values. The device developed had good agreement with standards of sphygmomanometer. In this context, Opsens' OPP-M developed a cardiovascular blood pressure monitoring system using a fibre optic sensor. Dynamic temperature measurement was also made possible with this device, and less moisture was induced during the monitoring process. A reduction in size makes the device more favourable for invasive cases. Devices possessed immune to electromagnetic interference, magnetic resonance imaging, and any other invasive surgery instruments. The integration of the sensor was made easy with a cost-effective strategy. The optical fibre pressure sensor used an optical detector, light source, signal processing module, etc. It was observed that the blood pressure measurement based fibre optic system is reliable and faster in terms of sensitivity and quality factors. Most promising technologies and improvements carried under optical MEMS systems for blood pressure measurement applications used the concept of FBG system, fibre optic system, optical ring resonator, photonic crystal membrane, and interferometers (94, 95). Silica optical fibre-based FBG sensor was designed and discussed for strain measurement. Induced pressure by pulse waves from human arteries is the signal for the FBG sensor. Since silica optical fibre may undergo fatigue fractures, plastic optical fibre was introduced to resolve this issue. Pulse waveform from the human arteries was obtained. Based on the results obtained, the measurement of waveform plastic optical FBG was the better solution compared with others in terms of accuracy and sensitivity (96)[4]. In another work, Katayama et al. proposed wearable blood pressure measurement for tracing the sudden change in human blood pressure in the arteries. FBG optical pressure sensor was embedded in the textile or garments for the easy measure of pressure waves form in the human body. The partial least square method was followed for calibrations of pulse waveform and induced FBG pressure. The sensing accuracy obtained by this sensor was comparable to conventional sensors (82, 96, 97)[4].

Bladder Pressure Monitoring

We know that pressure is a critical parameter that is used to represent the state of all organs in the body. Hence the measure of pressure is very crucial in the medical field. The conventional system uses the fluid base wherein the pressure is indirectly transmitted from the required organ. These often have a delay in measurement since fluid lines are used as a medium of measurement. The traditional methods of measuring pressure have always proven to have minor errors. Usually, there are no sensors places in the body of a patient since it limits the movement of a patient. Hence, pressure is measured when the person is hospitalised or before and after a problem. Due to these factors, we do not have any constant measuring of pressure in the human body, and it does not reveal the underlying problems faced by people. In this context, Poeggel discussed the urodynamic analysis for evaluating the dysfunction of the lower urinary tract during the filling and voiding process of the bladder. Pressure measurement of the bladder during this process made the urologist provide a required treatment to a patient based on individual diseases. The technique was based on two optical pressure sensors and a temperature sensor. A real-time pressure measurement system was made possible by this method. To overcome this difficulty, we can design lightweight, small-sized, and efficient MEMS. MEMS can be used with minimal efforts using invasive techniques that can continuously monitor the condition of the patient all throughout. Urinary incontinence, which is the involuntary loss of urine from the bladder, occurs in ∼10% of the male population worldwide. This procedure was performed by a skilled professional. The procedure consisted of inserting a catheter into the urethra of a patient into the urinary bladder, where there is a transducer in one end. The pressure and the reference pressure were measured in the rectum through fluid lines. The target pressure of the urethra was compared with the reference pressure of the rectum and analysed. It has a substantial margin of error due to the movement and measurement errors, however, pressure is a vital parameter to measure the performance of organs. For effective diagnostic, high quality and high resolution in measurements is necessary. In organs such as the urinary bladder, water-filled lines have their limitations and have still been practised throughout the decades. We should make use of the advancement in photonic MEMS, which offers a very small pressure sensor with precision and accuracy. We could conclude that there are more limitations to the values measured with water-filled lines. Hence, an in-target pressure sensor from the organs is much advanced and reliable in detecting any changes. In the meantime, optical fibre measurement systems are also considered for monitoring the temperature. The system consists of an encapsulated fibre optic pressure sensor and a Fabry-Perot pressure sensor. The optical fibre used in this case is FBG pressure sensors. The optical fibre pressure measurement system is compatible with the

[4]*Opsen's Solution*. Available online at: https://opsens-solutions.com/ (accessed July 20, 2019).

clinical phase. The whole system is compared with the traditional biocompatible urodynamic measuring system (98).

Cochlear Implant Application

Deafness affects about 10% of the general population, and the rate increases to more than 30% for those aged over 65. However, a hearing prosthesis such as a cochlear implant can provide shearing and significantly improve the quality of life. It is the most important advancement in the treatment of hearing impairment. In another work, Yao developed a fibre optic-based bending sensor for cochlear implantation. It has proved a great success in providing implantation for deafness in children and adults. Implantation using a cochlear electrode array is a complex surgical process, and there is a chance of damage to the inner wall of the cochlear, i.e., scala tympani. Since the electrode has to bend at various angles, and if the electrode does not bend at the required angle, damage to the cochlear may be severe. To avoid this, researchers developed a bend sensor for implantation surgery (99). In this context, Balster has investigated hearing restoration by optical stimulation alternative to electric stimulation. Activation of the auditory system was made by light fibre insertion into the coiled cochlea. Light fibre force measurement was conducted, and results were compared with traditional electric stimulation study. The results showed that minimal painful implantation surgery optical stimulation was considered feasible (100). In another work, fibre optic-based intracochlear pressure measurement system was developed. A 81 μm-diameter superluminescent diode (SLED) and other modes of optical fibre were embedded inside the transducers. These cochlear fibre optic sensors were helpful for early monitoring of hearing loss, as well as the sound pressure measurement of temporal bones and mechanical stimulation of the intracochlear system (83, 101, 102).

The work of Starovoyt et al. discussed optical coherence tomography for intracochlear structure. Studying the temporal bone by using OCT was the main aim of the authors. The authors focused on obtaining high-resolution images using optical coherence tomography qualitative and quantitative characteristics of intracochlear structures. The importance of OCT images in intracochlear implant surgery was explored. Implantable sensors for hearing devices and wearing the external hearing aid causes discomfort for most people. Different designs of the implantable sensor were evaluated and compared (103–105). **Figure 10C** shows the view of the human cochlear system. Key applications of optical MEMS sensors are presented in **Table 2**.

Other Applications

The study of Samavati et al. discussed the development of the COVID-19 virus diagnosis kit using a novel Go- decorated Au/FBG sensor. The saliva of a patient is considered a favourable testing specimen for COVID-19. In this study, salivary samples was collected from different types of people. The FBG-based probe has been used to detect the COVID-19 virus. The FBG probe was immersed in healthy saliva and saliva with the COVID-19 virus. The FBG probe detected the presence of the virus. The density of the virus was 1.2×10^8 copies/ml in saliva. The maximum change in wavelength observed was 1.12 nm and an intensity change of 2.01 dB (106). In this context, Sharath U designed and developed different types of FBG devices for novel medical applications. The author designed an FBG pulse recorder for cardiology applications and individual bite force monitoring devices. Bite force in men and women has been differentiated with the help of this device (107). By considering the same concept, author Jha et al. proposed an FBG sensor-based glove to monitor the flexure of 10 finger joints. The author evaluated the accuracy and repeatability of all 10 sensors. The FBG-based glove has been designed and tested with an FBG interrogator. Stroke patients were benefitted from the device developed (108). Similarly, in another work carried by Alwin T and co-authors, FBG sensor was used as a multi ocular sensor. In this case, FBG was utilised for sensing temperature, pressure, and chemical concentration (109). The human respiration rate was monitored with FBG based sensor with an Arduino interface. The respiration rate was also calculated during coughing and handshaking. Compared with an electric sensor, FBG-based sensor has shown prominent results (110). In this context Tavares et al. embedded an FBG sensor in a wheelchair. This wheelchair was mainly used to monitor the ulcer in bone areas of the seat. The FBG sensor was also used for sensing dorsal flexion and extension movements. Spending the time in job for many hours with bad posture will lead to back musculoskeletal disorders. Thus, FBG sensing that was developed has shown promising results for future fabrication (111, 112).

In another work, the authors developed a system for monitoring soft finger deformation, which was driven by shape memory alloy with an FBG sensor embedded. Firstly the structural and mechanical properties of a finger were considered for analysis. The central wavelength for change in curvature of material due to shape memory alloy deformation was monitored with FBG sensors. Soft finger deformation was analysed with ABAQUS software for analysing the FEA simulations (113). In this context, Mo et al. designed optical sensing needles for tissue identification. Optical sensing needles were used for minimally invasive surgery. Optical sensors were mainly used because of their immune to magnetic resonance. Intensity-modulated Fabry-Perot interference was considered for optical needles. Developed optical sensing needles were successfully implemented in the medical field (114). Application of intravascular catheterization was used in minimal vascular surgery considered distal force sensing system. The designed sensor has shown higher repeatability and stability. With the distal tip force sensor, it was obtained from the result that the stability of each flexure movement could be controlled with the FBG-based sensor (115). Non-invasive methods to measure wrist angular handgrip devices had been developed to measure the handgrip strength of humans with muscular-skeletal disorders. Handgrip strength was measured in different angular positions. The FBG sensor was encapsulated in the mechanical package and its variation in strain with peak resonance wavelength was monitored for various ranges of applied pressure (116).

Lower back pain is one of the most common muscular-skeletal disorders affecting workers. Video terminal workers are the most affected people with lower back pain. Prompt detection of wrong

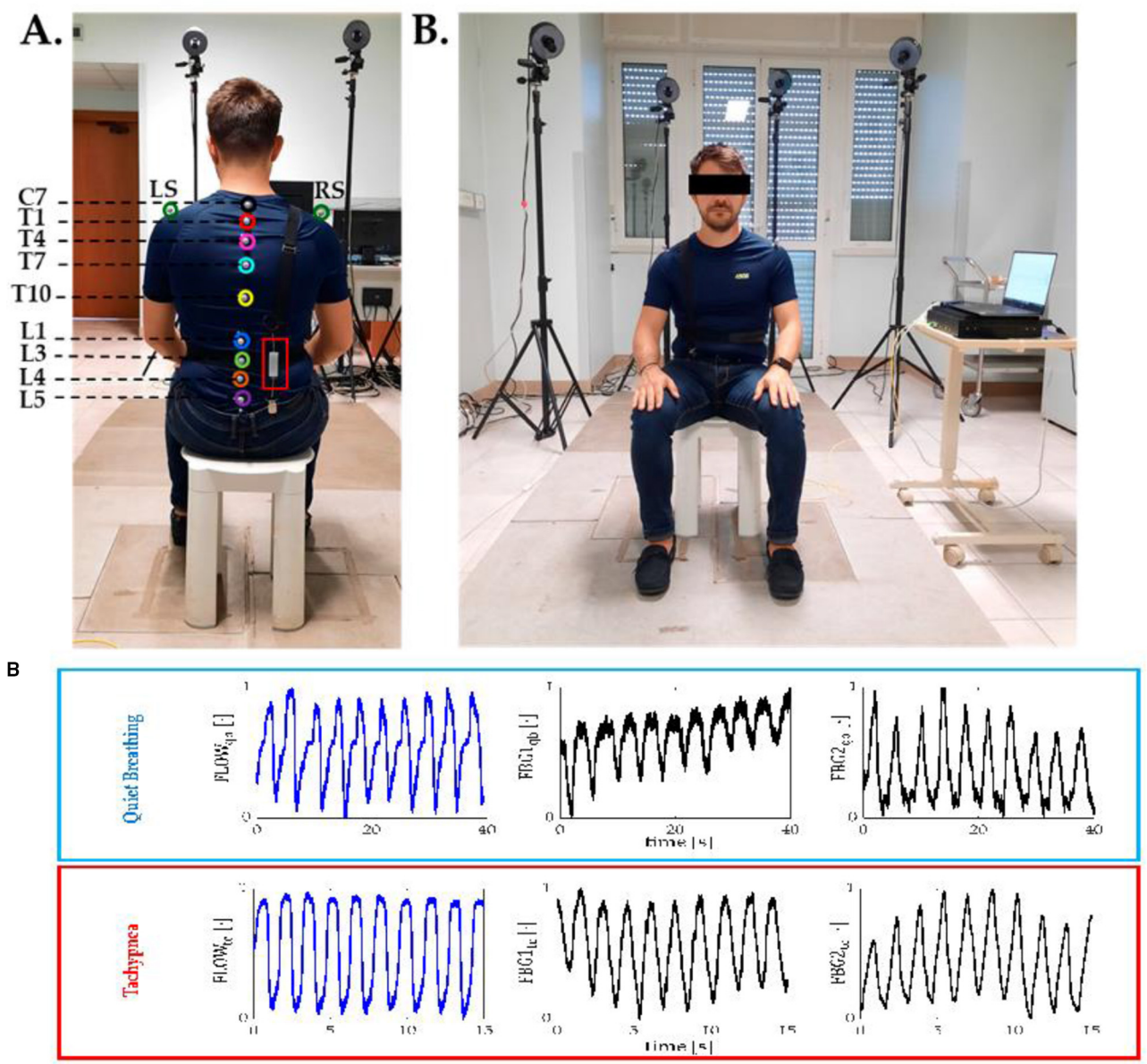


FIGURE 11 | (A) Back view showing the posterior part of the wearable devices with embedded sensor and and front view showing the anterior part of the embedded sensor (118). **(B)** Signal collected by flow meter and FBG during quite breathing and tachypnea (119).

seating postures is more critical. A lumbar movement monitoring device has been developed with FBG sensing technology. The sensitivity of the FBG device developed was 0.20 nm/mε^{-1}. This is the first device with silica embedded FBG sensing technology for lumbar motion monitoring for worker safety to overcome lower back pain. **Figure 11** shows the wearable device installed in person back for experimental purposes. The study of Lo Presti et al. considered an FBG sesor embedded in silica gel for monitoring neck movements and respiratory frequency. Neck movements during flexion-extension were investigated. In the meantime, FBG sensors were embedded in the human chest and monitored during quiet breathing and tachypnea (61, 117). **Figure 11A** shows the different views of a person with a wearable device for experimentation.

In the same context, FBG sensors are used in human gait analysis, including diagnosis, rehabilitation, and sports. The FBG sensor encapsulated in the mechanical package was used to monitor the extension and flexion movement. This evaluation is useful in monitoring walking disorders and pathologies. FBG wearable system was developed with flexible polymer matrix is

TABLE 3 | Summary of application.

Sl. No	Application	Summary of results
1	Dental	❖ Prediction of dental carries using optical sensor ❖ Incorporation of FBF sensor for dental carries ❖ Strain pattern evaluates for different parts of dental region ❖ Evaluation of dental filler material with FBG ❖ Evaluation of biting force with FBG sensor
2	Intervertebral disc	❖ Load distribution over intertribal disc during tensile and compressive stress has been monitored. ❖ Incorporation of FBG sensor spinal cord in different sitting position
3	Lumbar motion	❖ Load distribution over lumbar disc during tensile and compressive stress has been monitored. ❖ Incorporation of FBG sensor in lumbar position
4	Joint angle, gait analysis, insole pressure	❖ Integration of FBG sensor in insole for foot pressure monitoring in diabetic and arthritis patients. ❖ Embedding the sensors in insole and checking pressure distribution for different types of foot and in different experimental conditions such as walking, standing, climbing, age, and sex and in different patients.
5	Communication assistance	❖ Development of communication assistance based on eye blinking integrated with FBG sensor signals
6	Pulse wave and respiratory, body pressure	❖ Different pressure points of the body are monitored with FBG sensor ❖ Respiratory pressure is monitored with FBG sensor using wrapping the FBG sensor based belt around chest
7	Intraocular	❖ Optical fibre based sensor has been used to detect the glaucoma in patients
8	Urodynamic measurement	❖ Optical fibre has been incorporated with catheter to
9	Microfluidic system	❖ FBG sensor has been integrated with different microfluidic channel for biosensing applications
10	Handgrip	❖ FBG sensor for integrated with gloves, designed handgrip models,

embedded with FBG arrays as knee embrace. The developed FBG wearable system has shown sensitivity up to 3.944 nm-mm^{-1} (120). **Figure 11B** shows the signal collected by the flow metre and FBG during experimentation. In another work, Ran et al. proposed an FBG-based immunosensor for cardiac biomarkers. The concept of temperature cross-sensitivity was considered during the development of proof of concept. The tool was used to identifying acute heart diseases. In a similar case, FBG sensors were embedded in T-shirts to monitor the respiratory rate during cycling. Massaroni et al. proposed an FBG sensing system embedded in cloth for monitoring human body breath monitoring. Twelve FBG based sensors were embedded in the T-shirts. Four volunteers were considered during the study (49, 118, 119, 121–125)[5]. FBG sensors are also used to measure the pose of different medical instruments used in surgery, wherein FBG sensors are embedded in catheters. Flexible movement of catheters during surgery is monitored with an FBG sensing system (126–128). **Table 3** shows the summary of the application. **Figure 12** shows the major application of optical MEMS.

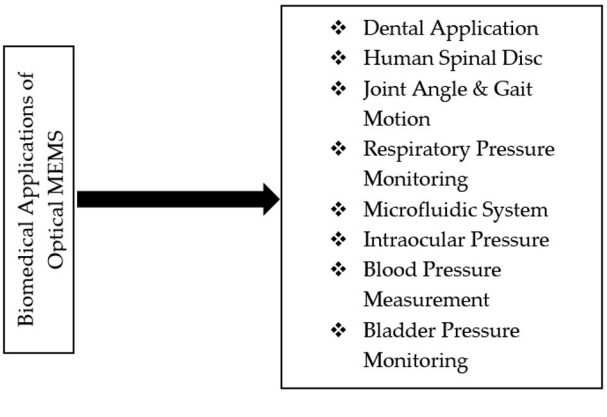


FIGURE 12 | Major applications of optical MEMS in biomedical.

KEY IMPACTS AND THE CHALLENGES OF THE OPTICAL MEMS FOR THE BIOMEDICAL APPLICATION

The authors investigated crucial factors of the optical MEMS sensing system by considering many research papers and concluded three key elements. These key factors are based on optical MEMS devices, optical MEMS applications, and current optical MEMS technology trends. A detailed analysis of these critical factors is grouped in **Table 4**. Conventional MEMS sensor uses piezoelectric or shapes memory alloy, transducers to convert physical phenomena and into electrical signals; these signals are scaled, conditioned, and digitised to the required applications. Despite their advantages, these sensors are impractical to use in all applications. Instead, the optical MEMS sensor offers excellent solutions to many challenges faced by MEMS sensors. Sometimes, the optical MEMS sensor works similarly to the MEMS sensor but uses light instead of electricity and silica glass instead of copper. It was observed from the literature that most biomedical applications use optical fibre or FBG technologies for monitoring pressure, temperature, and many other physical parameters. Different optical MEMS systems, such as photonic integrated microcantilevers, MZI, and ring resonators, can penetrate as emerging technology until laboratory testing costs and complexity and reliability issues of these

[5] *Distributed and Single Point Fibre Optic Sensing Systems Forecast*. Available online at: http://www.igigroup.com/ (accessed July 20, 2019).

TABLE 4 | Critical factors of optical MEMS.

Key factors for the optical MEMS methods/designs	References	Total studies
Key factor 1	(1–49)	49 papers
Key factor 2	(50–104, 129)[12345]	60 papers
Key factor 3	(49, 118, 119, 123–128)	28 papers

sensing systems are resolved. Critical success factors of the optical MEMS till 2020 include evaluation of sensor performance such as its sensitivity, Q factor, detection limit, Optimization of various sensing structures, etc. **Table 4** shows that research work carried out extensively with optical sensing systems such as photonic crystal structures, MZIs, and optical microcantilevers. Performance enhancement of various optical MEMS sensing structures has been carried out. It was possible to successfully implement the FBG and an optical fibre sensing system in many real-time applications. Still, authors are provided scope for improvement in much research work. Feasible sensitivity and detection limit is achieved based on a specific application. Most of the sensors were designed and analysed with transverse electric mode. It is noted that real-time implementation of other sensing systems apart from FBG and optical fibre sensor requires plenty of sensing system characterisation. **Figure 13** shows major applications of optical MEMS in biomedical.

Three critical factors are identified in the survey of optical MEMS for biomedical applications, and are clearly distinguished as Key factors 1, 2, and 3. Research articles on Key factor 1 highlighted the issue of pre-fabrication stages, such as design and simulations. Performance evaluation of sensors in terms of sensitivity, detection limit by design optimization, modified refractive index, and different material assignment for the sensing structure is discussed in Key factor 1. Applications in which optical MEMS sensors have been successfully implemented is considered as Key factor 2. From Key factor 2, it is observed that there is substantial usage of optical fibre sensors, FBG, and MEMS-based optical coherence tomography, optical MEMS actuators, etc. survey on the present market status of optical MEMS are discussed in Key factor 3. The optical MEMS design and testing phase faces technical challenges in the presence of mechanical elements during the integration process involving flow, pressure, movement, and analysing temperature. The combination of the multi-function device makes the design and testing complex during the development process of the optical MEMS sensor. Most organisations have no access to the fabrication facility, even if there is interest in exploring the technologies. Packaging can be expensive and knowledge of fabrication challenging aspect in the case of optical MEMS.

An overview of optical MEMS market growth in 2019 offers a complete analysis for better understanding the completion in the market and implementing the same in research activities. The optical MEMS market fragmented in different major companies like S.T. Microelectronics, Analogue Devices, Freescale Semiconductor, Texas Instruments, Boston Micromachines, and Memscap. As shown in **Figures 14A,B** below, the maximum market size of optical MEMS in the current era in the field of Biomedical Engineering is preceded by Defence and Security and Smart structures. The least volume of the market is observed in the small structures. Under the optical MEMS category, the original product is the projection system, microbolometers, MZI, optical cantilevers, ring resonators, and FBG sensors.

The fibre optic sensor market is projected in 2017 to be $1,008 million in 2021 and reaching above $1,033 million in 2023. Most commercial efforts focused on medical and military applications. Due to the continuous improvement in the FBG writing system, there is an increase in demand. Among the different types of FBG types, uniform FBG type is expected to have compound annual growth rate over 14% as forecast in upcoming days. Photonic crystal biosensor, optical ring resonators type of sensing system has shown less trend presently due to sophisticated experimental setup and more importantly because of the technology readiness level of photonic sensing devices. Most of the photonic crystal sensor has to be cascaded which leads to the difficulty over fabrication. Bringing out the photonic sensing system in the main market, except for fibre optics-based sensors, takes few years of time due to this challenging process involved. The global market of photonic crystals has reached up to $43.119 billion during 2015 and it will be $60.230 million during 2022. Photonic crystal sensing research has got attention from other academicians, as well as from scientists. Research filed such as industrial telecommunications, biomedical field, LED display, and optical sensing technology has shown remarkable growth with photonic crystal from the past few years.

Obstacles for the growth of the optical MEMS device mainly include design complexity, reliability, fabrication access, packaging facility, and knowledge of fabrication. The optical output parameters such as Q-factor, sensitivity, and transmission efficiency mainly depend upon the final packaging of the device. Since the optical MEMS technology involves more moving parts controlling the device, accuracy requires expertise in a multidisciplinary field. Despite these challenges in optical MEMS technology, it overcomes many disadvantages faced by MEMS technology, such as high sensitivity, less power consumption, EMI Insensitivity, and lightwave manipulation can be more comfortable with moving mechanical elements. The device-level design, fabrication, or packaging complexity is due to less exposure to optical MEMS technologies. Simplifying the device design and packaging process is the scope in the optical MEMS technologies. **Figures 14A,B** shows the curret trends of optic MEMS sensor in market.

Simplifying moving micromechanical parts in terms of size is essential for increasing the efficiency and performance of optical MEMS. Design and optimization of moving mechanical elements essential criteria to build the integrated optical sensing system. Using the sophisticated fabrication lab software tools for pre-checking the final model and packaging for reliability, the accuracy of readings is the primary solution to speed up the final fabrication to bring the device technology to commercial status. This technique mainly applies to the

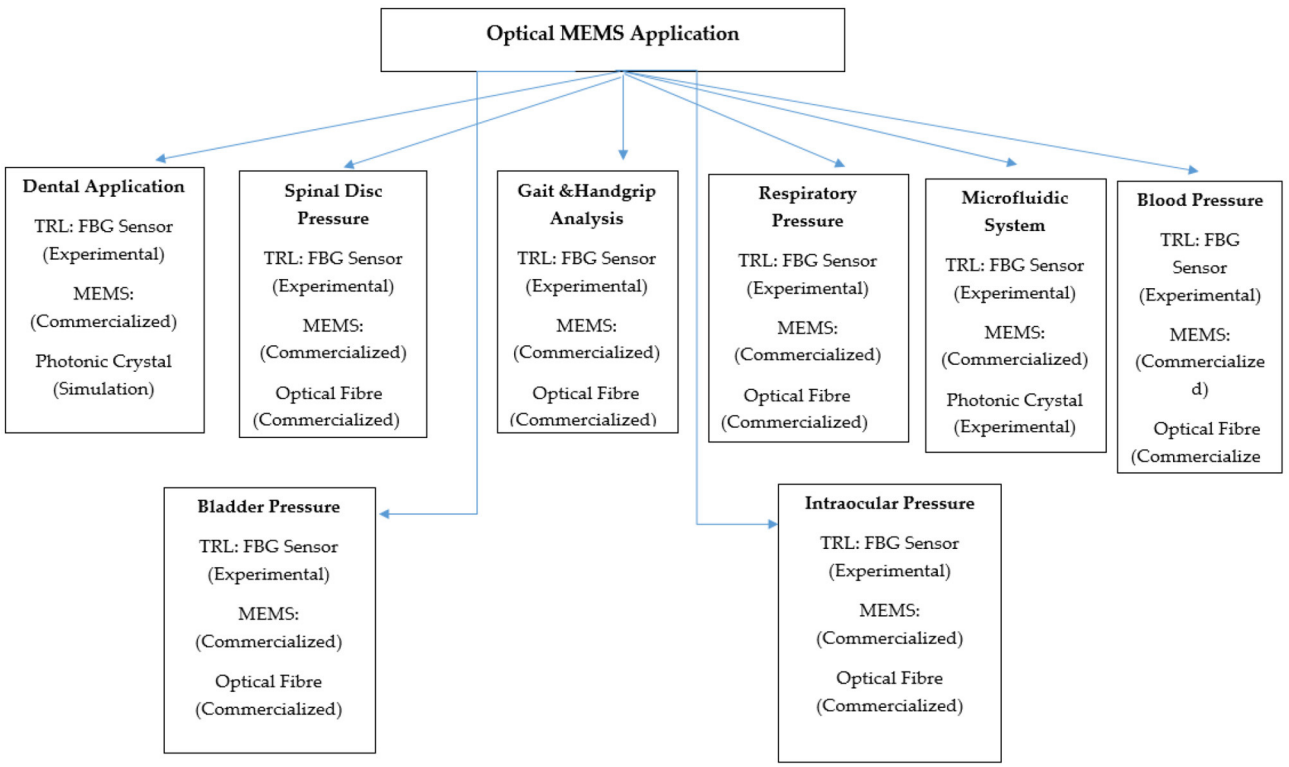


FIGURE 13 | Major applications and TRL (technology readiness level) of optical MEMS sensing system in biomedical.

photonic sensing system in the technology readiness level as experimental such as photonic crystal, photonic ring resonators, and MZI. In the case of FBG based sensing system, miniaturisation of FBG interrogator device in the form of chip will increase the demand in the market. Accurate debugging and design tuning, preparing virtual prototypes, Gathering practical feedback at the time of designing are the factors making up the successful implementation of the product. Designing the photonic integrated circuit with photonic crystals sensing technology and optical resonators or MZI is always challenging; actual implementations are possible with 3D emulators. This process shows how the microchip will work in 3D visualisation and helps the final working of the product after packaging. There will be unpredictable and immeasurable errors that have to be overcome for ease of fabrication. High tolerance needs to be maintained in the design for achieving targeted output. Coupling problem, temperature influence, controlling size, and position of holes are critical parameters affecting sensor performance due to the requirement of nanometer precisions scale.

SCOPE OF FUTURE WORK

The focus of this review was to explore the recent advancement of optical MEMS sensors and their application in the biomedical field. Key and important issues with optical MEMS had been identified. Using high sensitive MOEMS sensor for the harsh environment have operational and stability issues. There is always scope for developing sophisticated MOEMS sensing structures by analytical and simulation models to improvise the sensitivity, stability, and printable or packaging issues. Optomechanical sensors with high sensitivity could be developed by considering different mechanical properties, stiffness, the mass of sensing material, etc., thereby increasing response frequency for sensing pressure and displacement. The operational stability of the MOEMS sensor could also be increased by the coating of the sensing layer with different materials. There is always scope to investigate the integration of optical sensing layers with composite and organic materials to make packaging easier. Even though the FBG sensing system has reached commercialisation, the miniaturisation of the FBG system such as interrogator is the requirement of the present scenario to bring compact devices in the market.

- ❖ Variety of Photonic MEMS sensing systems have been considered in the literature, but minimal sensing system is implemented in biomedical applications due to sophisticated fabrication feasibility.
- ❖ There is a lot of scope for optimization of photonic MEMS sensing structure for implementation in biomedical applications since most of the photonic MEMS system, apart from optical fibre, is implemented in telecommunication.
- ❖ Fibre Bragg Grating is a mature technology and has plenty of scope for incorporating in biomedical applications

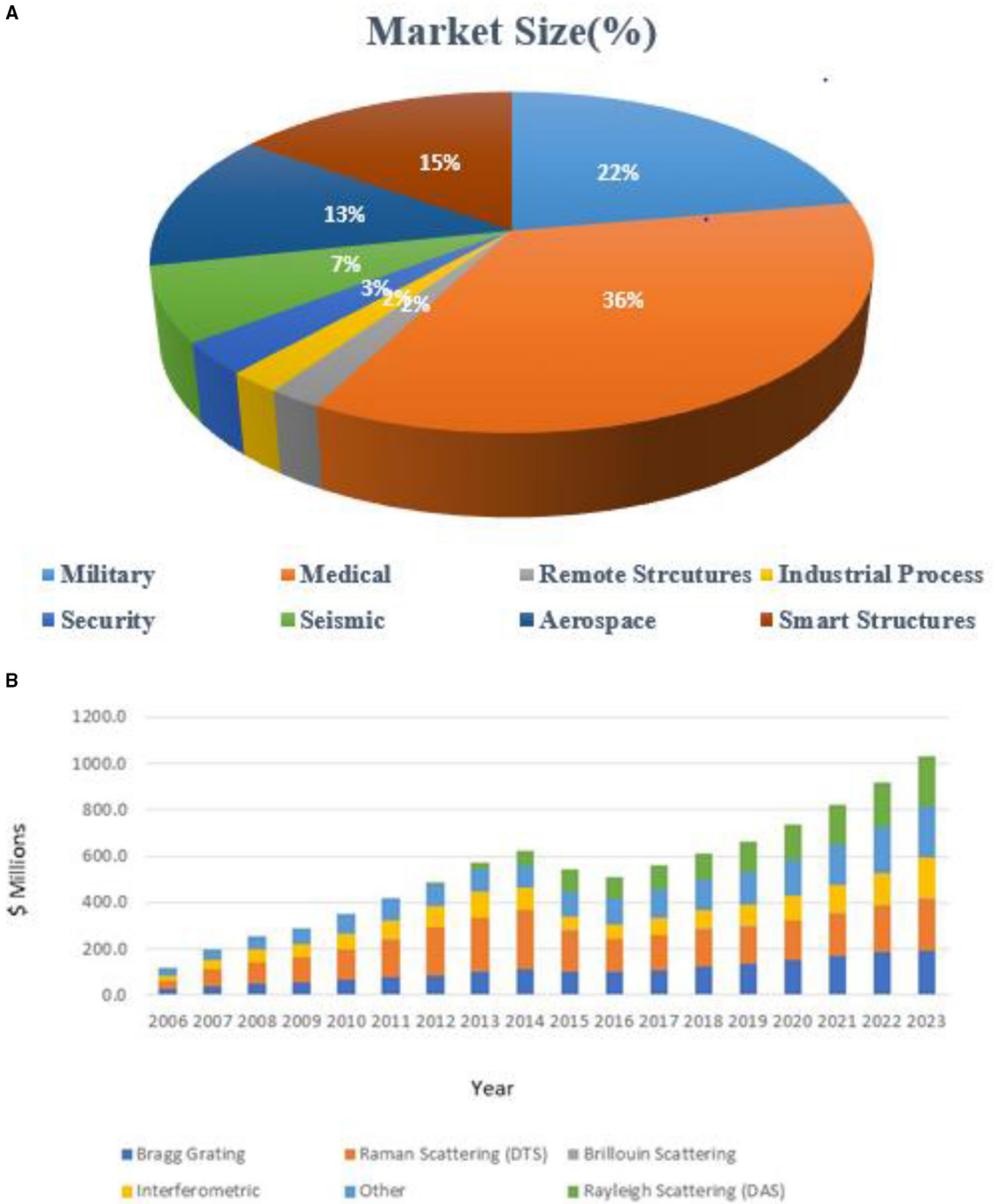


FIGURE 14 | (A) Presents trend of Optical MEMS market (127). **(B)** Optical MEMS market growth until 2023 (127–133).

to increase the accuracy, sensitivity, and durability of medical instruments.

- ❖ Recently published articles show possible future work in terms of design, different experimental conditions for specific applications.
- ❖ The obstacles to overcome soon include collecting experimental data, fabrication knowledge, and lack of fabrication facility.
- ❖ Dynamic analysis of each moving mechanical element is necessary before integrating these elements into the optical system to obtain more accurate results.
- ❖ Design and packaging, testing, sensing, latching, controllability, and reliability are the main challenges in developing optical mems sensors.

❖ Optical MEMS helps accelerate the deployment of different sensing systems in the biomedical field and other structural health monitoring fields.

CONCLUSION

This article reviewed different optical MEMS sensors and their application in the various biomedical fields with future trends and growth. To assist clinicians in knowing about the recent developments under optical MEMS, a clear possible visualisation was provided to implement the same technologies in the biomedical field. This paper has investigated the necessary information about optical MEMS, and its possible applications in the biomedical field have been assembled and discussed in detail. With many research articles under microcantilever-based optical MEMS diaphragm, interferometers under different forms were limited to simulation stages, but still showed possible future improvement towards improvising sensitivity with novel structure with experimental setups. Photonic crystal sensing and optical ring resonators had shown more discussion on developing mathematical modelling and numerical simulations. Structural optimization in terms of material properties, geometrical changes, and possible applications has been considered. Experimental techniques were considered in the field of photonic crystal, ring resonator, and MZI technologies. However, readers who are interested in readings of these photonic technologies would find plenty of unique design properties assigned during the structural optimization. Designers had increased their horizons towards finding new solutions for ease fabrication by simple designs with practical outputs. In the near future, photonic sensing technologies will control, integrate the network, and find real-time measurements in different biosensing applications. From the literature, it was also observed that most biomedical applications discussed here use the non-invasive method of parameter monitoring. There is always a possibility of the improvement of non-invasive clinical diagnosing with the optical sensor. Irrespective of complexity in fabrication, optical MEMS has consistently shown high operation speed, high optical efficiency, optical precision, reliability, readability, and scalability. Significant properties the optical MEMS material should incorporate are high stiffness, high fracture strength, fracture toughness, temperature, and chemical inertness. It was also found that optical coherence tomography major equipment used in biomedical image study. Although researchers had understood optical MEMS and experimented with a different application, implementation of the same in reality has not been possible in many cases due to sophisticated technology, lack of experimental setups, and the requirement of testing its consistency in performance. Since there are various fibre optic FBGs sensors, choosing the appropriate coating material for a suitable application is critical. As per the market survey, it was seen that there will always be an increase in demand for optical MEMS sensors with technical advancement.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AU, MH, MS, PS, and SI: writing, methodology, review, supervision, and project management. MH, RH, SI, SA-K, AS, and NV: review and funding. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been collaborated and supported by the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia under the Grant GGPM 2020 028, Taif University (TURSP- 2020/154), Saudi Arabia, Amity University, India, and the Oxford College of Engineering, India. And also this research was supported by the research grant of The National University of Malaysia (UKM) under the Grant GGPM 2020 028.

REFERENCES

1. Wibool P, Qiu Z. *Optical MEMS*. London: Intech Open (2012).
2. Motamedi M. *MOEMS: Micro-Opto-Electro-Mechanical Systems*. SPIE (2015) p. 636–8.
3. Ding L, Zeng ZC. In-fibre mach-zehnder interferometer based on three-core fibre for measurement of directional bending. *Sensors*. (2019) 19:205. doi: 10.3390/s19010205
4. Arokiaraj J, Vicknesh S, Ramam A. Integration of indium phosphide thin film structures on silicon substrates by direct wafer bonding. *J Phys*. (2006) 34:66. doi: 10.1088/1742-6596/34/1/066
5. Jindal SK, Raghuwanshi S. Realization of MOEMS pressure sensor using Mach Zehnder interferometer. *J Mech Sci Technol*. (2015) 29:3831. doi: 10.1007/s12206-015-0829-z
6. Lin B, Yi Y, Cao Y. A polymer asymmetric mach-zehnder interferometer sensor model based on electrode thermal writing waveguide technology. *Micromachines*. (2019) 10:100628. doi: 10.3390/mi10100628
7. Van-Pham D, Nguyen M, Nakanishi H, Norisuye T, Miyata Q. *Applications of Mach-Zehnder Interferometry to Studies on Local Deformation of Polymers Under Photocuring*. Lodnon: Intech Open (2017). doi: 10.5772/66955
8. Patel P, Hamde S. Analysis of sensitivity of optical mach zehnder interferometer (OMZI) sensor platform using rib waveguides. *Int J Innov Res Electr Electron Instrument Control Eng*. (2018) 6:6126. doi: 10.17148/IJIREEICE.2018.6126
9. Yadav M, Hovik J, Aksnes A. Highly sensitive lab-on-a-chip biosensor utilizing phase-modulated Mach-Zehnder interferometer. *Adv Photon*. (2018). doi: 10.1364/SENSORS.2018.SeTu4E.5
10. Wang B, Ni K. Improving the sensitivity of humidity sensor based on mach-zehnder interferometer coated with a methylcellulose. *Int J Opt*. (2018) 2018:5. doi: 10.1155/2018/5680128
11. Ouyang B, Li Y, Kruidhof M, Horsten R, Baets R, Koen W, et al. Highly sensitive silicon Mach-Zehnder interferometer-based ultrasound sensor. In: *SPIE, Proceedings Integrated Optics: Devices, Materials, and Technologies* San Francisco, CA. Vol. 11283 (2020). doi: 10.1117/12.2539573

12. Gao S, Ji C, Ning Q, Chen W, Li J. High-sensitive Mach-Zehnder interferometric temperature fibre-optic sensor based on core-offset splicing technique. *Optical Fiber Technol.* (2020) 56:102202. doi: 10.1016/j.yofte.2020.102202
13. Luan S, Enxiao D, *Improving the performance of silicon photonic optical resonator-based sensors for biomedical applications*. In: Library Open Collections (PhD Thesis), University of British Colombia, Vancouver, BC, Canada (2020). Available online at: https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0389559 (accessed October 21, 2021).
14. Wang R, Zhao J, Sun Y, Yu H, Zhou N, Zhang H, et al. Wearable respiration monitoring using an in-line few-mode fibre Mach-Zehnder interferometric sensor. *Biomed Opt Express*. (2019). 11:316–29. doi: 10.1364/BOE.376782
15. Ahsani V. *Design and analysis of fibre-optic Mach-Zehnder interferometers for highly sensitive refractive index measurement*. In: *UVicSpace* (PhD Thesis), University of Victoria, Victoria, BC, Canada (2020). Available online at: http://hdl.handle.net/1828/11724 (accessed October 21, 2021).
16. Zunic M. Design of a micro-opto-mechanical ultrasound sensor for photoacoustic imaging. In: *21st International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*. Cracow (2020). p. 1–8. doi: 10.1109/EuroSimE48426.2020.9152628
17. Hammadi A, Mahmood A. New optical fibre biosensor method for glucose in serum. In: *IOP Conference Series: Materials Science and Engineering, Volume 745, The Fourth Scientific Conference for Engineering and Postgraduate Research*. Baghdad (2019). doi: 10.1088/1757-899X/745/1/012049
18. Ahmed A, Hassan R, Othman NE. Improving security for IPv6 neighbor discovery. In: *International Conference on Electrical Engineering and Informatics (ICEEI)*. Denpasar (2015). p. 271–4. doi: 10.1109/ICEEI.2015.7352509
19. Rosminazuin R, Nurashikin A, Noreha M, Badariah B. and Burhanuddin Y. Fabrication of monolithic wheatstone bridge circuit for piezoresistive microcantilever sensor. In: *Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS* Montpellier. (2015). p. 1–5.
20. Rahim RA, Abubakkar S, Bais B, Majlis BY. Analysis of electrical responses of MEMS piezoresistive microcantilever. In: *Proceedings - 5th International Conference on Computer and Communication Engineering: Emerging Technologies via Comp-Unication Convergence, ICCCE*. (2014). p. 123–6. doi: 10.1109/ICCCE.2014.45
21. Mardhiah M, Badariah B, Norazreen Abd A, Rosminazuin AbR, Burhanuddin YM. Induced mass change technique for glucose detection in microcantilever-based sensors. In: *Proceedings IEEE International Conference on Semiconductor Electronics (ICSE2012)*. (2012).
22. Rosminazuin R, Badariah B, Burhanuddin M. Microcantilever release process using micromachining technology. *J Sci Technol Trop*. (2011) 7:105–8. Available online at: http://www.costam.org.my/publications/Vol%207%20No%202%20Dec%202011.pdf (accessed october 21, 2021).
23. Sang S, Zhang W, Zhao, Y. *Review on the Design Art of Biosensors*. London: Intech Open (2012). doi: 10.5772/52257
24. Lee, C, Thillaigovindan J, Chen C. Si Nano photonics-based cantilever sensor. *Appl Phys Lett.* (2008) 93:113113. doi: 10.1063/1.2987515
25. Xiang W, Lee C. Nanophotonics sensor based on microcantilever for chemical analysis. *IEEE J Select Top Quant Electron.* (2009) 15:1323–6. doi: 10.1109/JSTQE.2009.2016578
26. Cagri A, Savran W, Sparks A. Fabrication and characterization of a micromechanical sensor for differential detection of nanoscale motions. *IEEE J Microelectromech Syst.* (2002) 11:1057. doi: 10.1109/JMEMS.2002.805057
27. Marcelo M, Werneck R, Allil C, Bessie A, Ribeiro FV. *Guide to Fibre Bragg Grating Sensors*. London: Intech Open (2013).
28. Junior L, Graciliano E, Franco A, Costa M, Karam L, Kalinowski H et al. Bulk Fill flow resin contraction using 3D finite element model and calibration by Fiber Bragg Grating measurement. *Comp Methods Biomech Biomed Eng*. (2020) 23:1127–37. doi: 10.1080/10255842.2020.1789607
29. Kreuzer M. *Strain Measurement With Fibre Bragg Grating Sensor*. HBM. Available online at: http://www.ae.metu.edu.tr/seminar/strain-gage/FBG.pdfS (accessed October 21, 2021).
30. Serene M, Babu R, Alex ZC. Sensitivity analysis of micro-mass optical MEMS sensor for biomedical IoT devices. In: *Internet of Things and Personalized Healthcare Systems*. Berlin: Springer, Briefs in Applied Sciences and Technology (2019). doi: 10.1007/978-981-13-0866-6_1
31. Alvarez M, Zinoviev K, Moreno M, Lechuga L. Cantilever biosensors. *Opt Biosens*. (2008) 7:419–52. doi: 10.1016/B978-044453125-4.50012-7
32. Steven G, Noyce RR, Vanfleet HG, Craighead RC. High surface-area carbon microcantilevers. *Nanoscale Adv*. (2019) 1:1148. doi: 10.1039/C8NA00101D
33. Mathew R, Ravi Sankar A. A review on surface stress-based miniaturized piezoresistive SU-8 polymeric cantilever sensors. *Nano-Micro Lett*. 10:35. doi: 10.1007/s40820-018-0189-1
34. Salve M, Mayuri D, Rewatkar R, Balpande S, Kalambe J. Design and sensitivity analysis of micro-cantilever based biosensor for tumor detection. *Sensor Lett*. (2018) 16:1–5. doi: 10.1166/sl.2019.3981
35. Saurabh A, Jitendra K, Vishnu P. Highly sensitive MOEMS integrated photonic crystal cavity resonator for nano-mechanical sensing. *Opt Commun*. (2020) 474:126150. doi: 10.1016/j.optcom.2020.126150
36. Rajasekar R, Robinson S. Nano-pressure and temperature sensor based on hexagonal photonic crystal ring resonator. *Plasmonics*. (2019) 14:3–15. doi: 10.1007/s11468-018-0771-x
37. Vijay S, Chaudharya D, Mishra R, Sharma S. Hybrid dual core photonic crystal fibre as hydrostatic pressure sensor. *Optik*. (2020) 210:164497. doi: 10.1016/j.ijleo.2020.164497
38. Rajib G, Ghosh K, Chakraborty R. High resolution wide range pressure sensor using hexagonal ring and micromachined cantilever tips on 2D silicon photonic crystal. *Opt Commun*. (2019) 431:93–100. doi: 10.1016/j.optcom.2018.09.016
39. Bonefacino J, Tam H, Glen T. Ultra-fast polymer optical fibre Bragg grating inscription for medical devices. *Light Sci Appl*. (2018) 7:17161. doi: 10.1038/lsa.2017.161
40. Khan F, Denasi A, Barrera D, Madrigal J, Sales S, Misra S. Multi-core optical fibres with Bragg gratings as shape sensor for flexible medical instruments. *IEEE Sens J*. (2019) 2019:5878–84. doi: 10.1109/JSEN.2019.2905010
41. Jasjot K, Sahota GN, Dhawan D. Fibre Bragg grating sensors for monitoring of physical parameters: a comprehensive review. *Opt Eng*. (2020) 59:060901. doi: 10.1117/1.OE.59.6.060901
42. Zhenga C, Jin F, Zhao Y, Zheng M, Liu L, Dong X, et al. Light-driven micron-scale 3D hydrogel actuator produced by two-photon polymerization micro fabrication. *Sens Actuat B Chem*. (2020) 304:127345. doi: 10.1016/j.snb.2019.127345
43. Liu T, Pagliano F, van Veldhoven R. Integrated nano-optomechanical displacement sensor with ultra-wide optical bandwidth. *Nat Common*. (2020) 11:2407. doi: 10.1038/s41467-020-18579-2
44. Nguyen L, Schartner E, Otten D, Yu Z, Lancaster D, Heidepriem E, et al. Multi-point optical fibre pressure sensor. In: *Proceedings Volume 11200, AOS Australian Conference on Optical Fibre Technology (ACOFT) and Australian Conference on Optics, Lasers, and Spectroscopy (ACOLS)*. Washington, DC (2019). p. 112000X. doi: 10.1117/12.2541295
45. Zaki W, Correia R, Korposh S, Barrier R, Gill H, Morgan SP. Development of wearable optical-based fibre sensor system for pulse transit time measurement. In: *AIP Conference Proceedings*. Malaysia (2020) 2203:020013. doi: 10.1063/1.5142105
46. Yen Y, Chiu C. A CMOS MEMS-based membrane-bridge nanomechanical sensor for small molecule detection. *Sci Rep*. (2020) 10:2931. doi: 10.1038/s41598-020-60057-8
47. Raghuwanshi S, Singh S, Shadab A, Pandey P, Prakash O, Saini P. Sensitivity analysis of TiO2 coated fibre Bragg grating sensor for far infrared detection of chemicals. In: *Proceedings Volume 11233, Optical Fibres and Sensors for Medical Diagnostics and Treatment Applications*. (2020). p. 1123306. doi: 10.1117/12.2542908
48. Lv C, Wang S, Shi C. A high-precision and miniature fibre bragg grating-based force sensor for tissue palpation during minimally invasive surgery. *Ann Biomed Eng*. (2020) 48:669–81. doi: 10.1007/s10439-019-02388-w
49. Khan A, Donder S, Galvan F, Baena Y, Misra S. Pose measurement of flexible medical instruments using fibre bragg gratings in multi-core fiber. *IEEE Sens J*. (2020) 2020:2993452. doi: 10.1109/JSEN.2020.2993452

50. Pyo S, Lim Y, Lee W, Lee J. Study on application to the field of dentistry using optical coherence tomography (OCT). *J Korean Acad Prosthodont*. (2017) 55:100–10. doi: 10.4047/jkap.2017.55.1.100
51. Scandelari M, Carlos J, Silva C, Martelli C, Kalinowski H. Force monitoring in a maxilla model and dentition using optical fibre bragg gratings. *Sensors*. (2012) 12:11957–65. doi: 10.3390/s120911957
52. Hsiao T, Lee S, Sun C. Optical polarimetric detection for dental hard tissue diseases characterization. *Sensors*. (2019) 19:4971. doi: 10.3390/s19224971
53. Logozzo G, Franceschini A, Kilpelae M, Caponi L, Governi L, Blois A. Comparative analysis of intraoral 3d digital scanners for restorative dentistry. *Internet J Med Technol.* (2008) 2008:1. doi: 10.5580/1b90
54. Cheng R, Shao J, Gao X, Tao C, Ge J, Liu X. Noninvasive assessment of early dental lesion using a dual-contrast photoacoustic tomography. *Sci Rep.* (2016) 2016:21798. doi: 10.1038/srep21798
55. Li Z, Holamoge YV, Li Z, Zaid W, Osborn ML, Ramos A, et al. Detection and analysis of enamel cracks by ICG-NIR fluorescence dental imaging. *Ann NY Acad Sci*. (2020) 1475:52–63. doi: 10.1111/nyas.14374
56. Baysan A, Sleibi A, Ozel B. The quantification of surface roughness on root caries using Noncontact Optical Profilometry—an *in vitro* study. *Laser Dent Sci.* (2018) 2:229–37. doi: 10.1007/s41547-018-0041-4
57. Umesh S, Padma T, Srinivas SA. Fibre bragg grating goniometer for joint angle measurement. *IEEE Sens J.* (2018) 18:216–22. doi: 10.1109/JSEN.2017.2770176
58. Newell N, Little J, Christou A, Adams M, Masouros S. Biomechanics of the human intervertebral disc: a review of testing techniques and results. *J Mech Behav Biomed Mater.* (2017) 69:420–34. doi: 10.1016/j.jmbbm.2017.01.037
59. Dennison C, Dvorak M, Wilson D, Cripton P. Validation of a novel minimally invasive intervertebral disc pressure sensor utilizing in-fibre Bragg gratings in a porcine model: an *ex vivo* study. *Spine*. (2008) 33:E589–94. doi: 10.1097/BRS.0b013e31817c55e2
60. Yamaji T, Yamamoto K, Ootaka H, Bessho Y, Kobayashi Y, Ono R. Wearable lumbar-motion monitoring device with stretchable strain sensors. *J Sens.* (2018) 2018:7480528. doi: 10.1155/2018/7480528
61. Zaltieri M, Massaroni C, Presti D, Bravi B, Sabbadini R, Miccinilli S, et al. A wearable device based on a fibre bragg grating sensor for low back movements monitoring. *Sensors*. (2020) 20:E3825. doi: 10.3390/s20143825
62. Rezende C, Alves I, Marques MA, Naves SE. Polymer optical fibre goniometer: a new portable, low cost and reliable sensor for joint analysis. *Sensors*. (2018) 18:4293. doi: 10.3390/s18124293
63. Donno M, Palange E, Nicola F, Bucci G, Ciancetta C. New flexible optical fibre goniometer for dynamic angular measurements: application to human joint movement monitoring. *IEEE Trans Instrument Measur.* (2008) 57:925336. doi: 10.1109/TIM.2008.925336
64. Galvao J, Zamarreno C, Martelli C, Matias I. Smart carbon fibre transtibial prosthesis based on embedded fibre bragg gratings. *IEEE Sens J.* (2018) 18:1520–7. doi: 10.1109/JSEN.2017.2786661
65. Domingues F, Tavares C, Leitão C, Antunes P. Insole optical fibre Bragg grating sensors network for dynamic vertical force monitoring. *J Biomed Opt.* (2017) 22:91507–9, doi: 10.1117/1.JBO.22.9.091507
66. Vilarinho D, Theodosiou A, Leitão C, Marques A. POFBG-embedded cork insole for plantar pressure monitoring. *Sensors*. (2017) 17:2924. doi: 10.3390/s17122924
67. Lakho A, Yi-Fan Z, Hua J, Cheng-Yu H, Ahmed AZ. A smart insole for monitoring plantar pressure based on the fibre Bragg grating sensing technique. *Textile Res J.* (2019) 89:3433–46. doi: 10.1177/0040517519833977
68. Débora V, Theodosiou A, Maria F, Paulo A, Kyriacos K, Carlos M. Foot plantar pressure monitoring with CYTOP bragg gratings sensing system. *Biodevices*. (2018) 25–29. doi: 10.5220/0006533700250029
69. Monge J, Postolache O, Alexandre R, Domingues M, Antunes P, Viegas V. Fibre bragg gratings solution for gait assessment. In: *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. Dubrovnik (2020). p. 1–6. doi: 10.1109/I2MTC43012.2020.9128421
70. Padma S, Umesh S, Pant S, Asokan S. Fibre Bragg grating sensor-based communication assistance device. *J Biomed Opt.* (2016) 21:86012. doi: 10.1117/1.JBO.21.8.086012
71. Krehel M, Schmid M, Rossi R, Boesel L, Bona G, Scherer LJ. An optical fibre-based sensor for respiratory monitoring. *Sensors*. (2014) 14:13088–101. doi: 10.3390/s140713088
72. Nishiyama M, Watanabe M. Respiration rhythm monitoring in sleep based on weight movement using hetero-core fibre optic sensors. In: *ICCAS*. Gyeonggi-do (2010). doi: 10.1109/ICCAS.2010.5669885
73. Kam W, Mohammed MG, Leen O, Sullivan M, Keeffe S, Lewis S. All plastic optical fibre-based respiration monitoring sensor. In: *IEEE Sensors Conference proceedings, 2017, INSPEC Accession Number – 17434335:1–3*. Glasgow (2017). doi: 10.1109/ICSENS.2017.8233942
74. Massaroni C, Nicolò A, Lo Presti D, Sacchetti M, Silvestri S, Schena E. Contact-based methods for measuring respiratory rate. *Sensors*. (2019) 19:908. doi: 10.3390/s19040908
75. Nguyen T, Ichiki M. MEMS-based sensor for simultaneous measurement of pulse wave and respiration rate. *Sensors*. (2019) 19:4942. doi: 10.3390/s19224942
76. Arifin N, Agustina S, Irfan I, Dahlang T. Polymer optical fibre-based respiratory sensors: various designs and implementations. *J Sens.* (2019) 2019:6970708. doi: 10.1155/2019/6970708
77. Jiang H, Zou Y, Chen X. Continuous detection of micro-particles by fibre Bragg grating Fabry-Pérot flow cytometer. *Optic Expr.* (2018) 26:12579–84. doi: 10.1364/OE.26.012579
78. Cooksey G, Ahmed Z. *Optofluidic Temperature and Pressure Measurements with Fibre Bragg Gratings Embedded in Microfluidic Devices*. (2016). Available online at: https://arxiv.org/ftp/arxiv/papers/1603/1603.07743.pdf (accessed October 21, 2021).
79. Weizhi S, Mingzhu L, Changqing JY, Yanlin S. Direct-writing colloidal photonic crystal microfluidic chips by inkjet printing for label-free protein detection. *Lab Chip*. (2012) 12:3089–95.
80. Nunes P, Mortensen N, Mortensen JP, Klaus B. Refractive index sensor based on a 1D photonic crystal in a microfluidic channel. *Sensors.* (2010) 10:2348–58. doi: 10.3390/s100302348
81. Phuong T, Alex P, Nicolas W, Frank T. Development of a portable reader for an optical intraocular pressure sensor. In: *Conference: ASME 2019 28th Conference on Information Storage and Processing Systems*. San Diego, CA.
82. Haseda Y, Bonefacino J, Tam H, Chino S, Koyama S, Ishizawa H. Measurement of pulse wave signals and blood pressure by a plastic optical fibre FBG sensor. *Sensors*. (2019) 19:5088. doi: 10.3390/s19235088
83. Grossohmichen M, Salcher R, Püschel K, Lenarz T, Maier H. Differential intracochlear sound pressure measurements in human temporal bones with an off-the-shelf sensor. *Biomed Res Int.* (2016) 2016:6059479. doi: 10.1155/2016/6059479
84. Khoshnoud K, Silva C. Recent advances in MEMS sensor technology – biomedical applications. *IEEE Instrument Measur Magazine*. (2012) 15:8–14. doi: 10.1109/MIM.2012.6145254
85. Saccomandi P, Varalda A, Gassino R, Tosi D, Massaroni C, Caponero M, et al. Linearly chirped fibre Bragg grating response to thermal gradient: from bench tests to the real-time assessment during *in vivo* laser ablations of biological tissue. *J Biomed Opt.* (2017) 22:097002. doi: 10.1117/1.JBO.22.9.097002
86. Quack N. Silicon photonic MEMS: exploiting mechanics at the nanoscale to enhance photonic integrated circuits. In: *2019 Optical Fiber Communications Conference and Exhibition (OFC).* (2019). p. 1–3. Available online at: https://ieeexplore.ieee.org/document/8696652 (accessed October 21, 2021).
87. Gaber N, Altayyeb A, Soliman S, Sabry Y, Marty F, Bourouina T. On-channel integrated optofluidic pressure sensor with optically boosted sensitivity. *Sensors*. (2019) 19:944. doi: 10.3390/s19040944
88. Yanli Q, Rowe J, Mittal V, Banakar M, Yangbo W, Milos N, et al. Integration of mid-infrared SOI photonics with microfluidics. In: *Proc. SPIE 10923, Silicon Photonics XIV*. (2019). p. 109230I.
89. Daniel T, McCormick, Jung W, Ahn A, Chen Z. *Transducers 2007*. Lyon: International Solid-State Sensors, Actuators and Microsystems (2007).
90. Lee J, Park H, Du J, Balakrishna J, Sretavan D, Choo H. A microscale optical implant for continuous *in vivo* monitoring of intraocular pressure. *Microsyst Nanoeng.* (2017) 3:17057. doi: 10.1038/micronano.2017.57

91. Karrock T, Gerken M. Pressure sensor based on flexible photonic crystal membrane. *Biomed Optic Expr.* (2015) 6:12. doi: 10.1364/BOE.6.004901
92. Rezaie M, Gulari M, Franco F, Chronis N. Powerless optical micro sensor for monitoring intraocular pressure with keratoprosetheses. In: *Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS & EUROSENSORS XXVII), 2013, Transducers & Eurosensors XXVII: The 17th International Conference*. Barcelona (2013).
93. Samuel J, Haeri P, Jeong L, Hyuck C. Effect of optical aberrations on intraocular pressure measurements using a microscale optical implant in *ex vivo* rabbit eyes. *J Biomed Opt.* (2018) 23:047002. doi: 10.1117/1.JBO.23.4.047002
94. Min KP, Kim DY, Kim GW. Optical coherence tomography-based intraocular pressure monitoring principle. In: *Proceedings Volume 11378, Nano-, Bio-, Info-Tech Sensors, and 3D Systems IV*. (2020). p.113781A.
95. Katsuragawa Y, Ishizawa H. Non-invasive blood pressure measurement by pulse wave analysis using FBG sensor. In: *IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*. (2015). p. 511–5. doi: 10.1109/I2MTC.2015.7151320
96. Liu R, Correia R, Ballaji H, Korposh S, Gill S, Morgan S. Optical fibre-based pulse oximetry sensor with contact force detection. *Sensors*. (2018) 18:3632. doi: 10.3390/s18113632
97. Katayama K, Chino S, Koyama S, Keisaku F. Verification of blood pressure monitoring system using optical fibre sensor: —tracing sudden blood pressure changes. *J Fibre Sci Technol.* (2020) 76:79–87. doi: 10.2115/fiberst.2020-0008
98. Poeggel S, Duraibabu S, Tosi D, Leen G. Differential *in vivo* urodynamic measurement in a single thin catheter based on two optical fibre pressure sensors. *J Biomed Opt*. (2015) 20:037005. doi: 10.1117/1.JBO.20.3.037005
99. Li Y. Fibre-optic bending sensor for cochlear implantation. In: *Proceedings of the SPIE, 2006, 6047, I.D.—60473N*.
100. Balster S, Wenzel GI, Warnecke A, Reuter G. Optical cochlear implant: evaluation of insertion forces of optical fibres in a cochlear model and of traumata in human temporal bones. *Biomed Eng.* (2013) 59:1–10. doi: 10.1515/bmt-2013-0038
101. Olson E, Nakajima H. A family of fibre-optic based pressure sensors for intracochlear measurements. In: *Proceedings of SPIE 9303*. (2015). doi: 10.1117/12.2178056
102. Arora K. *Cochlear Implant Stimulation Rates and Speech Perception*. London: Intech Open (2012), doi: 10.5772/49992
103. Starovoyt A, Putzeys T, Wouters J, Verhaert N. High-resolution imaging of the human cochlea through the round window by means of optical coherence tomography. *Sci Rep*. (2019) 9:14271. doi: 10.1038/s41598-019-50727-7
104. Calero D, Paul S, Gesing A, Alves F, Cordioli J. A technical review and evaluation of implantable sensors for hearing devices. *Biomed Eng.* (2018) 17:23. doi: 10.1186/s12938-018-0454-z
105. Alexander D, Daniel K, Tobias M. Towards the optical cochlear implant: optogenetic approaches for hearing restoration. *EMBO Mol Med.* (2020) 12:e11618. doi: 10.15252/emmm.201911618
106. Alireza S. Sustainable and fast saliva-based COVID-19 virus diagnosis kit using a novelnGO-decorated Au/FBG sensor. *Chem Eng J.* (2020) 2020:127655. doi: 10.1016/j.cej.2020.127655
107. Sharath U. *Fibre Bragg Grating Based Sensing Devices for Novel Applications in Bio-medical and Engineering Field, electronic thesis and dissertation of IISc*. (2020). Available online at: https://etd.iisc.ac.in/handle/2005/4556 (accessed October 21, 2021).
108. Jha K, Gajapure K, Chakraborty L. Design and evaluation of an FBG sensor-based glove to simultaneously monitor flexure of ten finger joints. *IEEE Sens J.* (2021) 21:7620–30. doi: 10.1109/JSEN.2020.3046521
109. Meri A, Hasan MK, Safie N. Success factors affecting the healthcare professionals to utilize cloud computing services. *Asia-Pacific J Inform Technol Multimedia.* (2017) 6:31–42. doi: 10.17576/apjitm-2017-0602-04
110. Gautam AK, Priye V. FBG-based respiration rate sensing with arduino interface. *IEEE Sens J.* (2020) 20:9209–17. doi: 10.1109/JSEN.2020.2989004
111. Tavares C, Domingues M, Paixão T, Alberto N, Silva H, Antunes P. Wheelchair pressure ulcer prevention using FBG based sensing devices. *Sensors*. (2020) 20:212. doi: 10.3390/s20010212
112. Zaltieri Z. Feasibility assessment of an FBG-based wearable system for monitoring back dorsal flexion-extension in video terminal workers. In: *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. Dubrovnik (2020). p. 1–5. doi: 10.1109/I2MTC43012.2020.9128974
113. Yin H, Zhang Y, Wang J, Cao J. Position prediction of soft finger driven by SMA based on fibre bragg grating sensor. *IEEE Sens J.* (2021) 21:2951–62. doi: 10.1109/JSEN.2020.3022421
114. Mo Z, Li J, Xu W, Broderick NGR. The design of optical sensing needles for tactile sensing and tissue identification at needle tip. In: Billingsley J, Brett P, editors, *Mechatronics and Machine Vision in Practice 4*. Cham: Springer (2021).
115. Akinyemi T. Development of a Millinewton FBG-based distal force sensor for intravascular interventions. In: *16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. (2020). p. 511–5. doi: 10.1109/ICARCV50220.2020.9305395
116. Honnungar IA. Design and calibration of fiber bragg grating sensor for analysis of real time skeletal hand muscle strength. In: *IEEE International Conference for Innovation in Technology (INOCON)*. Bangalore (2020). p. 1–5.
117. Presti D, Carnevale A, D'Abbraccio J, Massari L, Massaroni C, Sabbadini R, et al. A multi-parametric wearable system to monitor neck movements and respiratory frequency of computer workers. *Sensors*. (2020) 20:536. doi: 10.3390/s20020536
118. Islam S, Khalifa OO, Hashim AHA, Hasan MK, Razzaque MA, Pandey B. Design and evaluation of a multihoming based mobility management scheme to support inter technology handoff in PNEMO. *Wireless Pers Commun.* (2020) 114:1133–53. doi: 10.1007/s11277-020-07412-0
119. Hasan M, Ismail F, Islam S, Hashim W, Ahmed MM, Memon I. A novel HGBBDSA-CTI approach for subcarrier allocation in heterogeneous network. *Telecommun Syst.* (2019) 70:245–62. doi: 10.1007/s11235-018-0473-x
120. Resta P. A wearable system for knee flexion/extension monitoring: design and assessment. In: *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*. (2020). p. 273–7. doi: 10.1109/MetroInd4.0IoT48571.2020.9138192
121. Ran Y, Long J, Xu Z, Yi Y, Deming HL, Yongkang Z, et al. Harmonic optical microfibre Bragg grating immunosensor for the accelerative test of cardiac biomarker (cTn-I). *Biosens Bioelectron.* (2021) 179:113081. doi: 10.1016/j.bios.2021.113081
122. Massaroni C, Nicolò P, Sacchett ES. Respiratory monitoring during cycling exercise: performance assessment of a smart T-shirt embedding fibre optic sensors. In: *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*. (2020). p. 49–53. doi: 10.1109/MetroInd4.0IoT48571.2020.9138307
123. Hasan M, Ahmed M, Hashim AHA, Razzaque A, Islam S, Pandey B. A novel artificial intelligence based timing synchronization scheme for smart grid applications. *Wireless Personal Commun.* (2020) 2020:1–18. doi: 10.1007/s11277-020-07408-w
124. Islam S, Abdalla AH, Hasan MK. Novel multihoming-based flow mobility scheme for proxy NEMO environment: a numerical approach to analyse handoff performance. *Scienceasia*. (2017) 43:27–34. doi: 10.2306/scienceasia1513-1874.2017.43S.027
125. Hasan MK, Yousoff SH, Ahmed MM, Hashim AHA, Ismail AF, Islam S. Phase offset analysis of asymmetric communications infrastructure in smart grid. *Elektronika ir Elektrotechnika*. (2019) 25:67–71. doi: 10.5755/j01.eie.25.2.23209
126. Dhanamjayulu C, Nizhal U, Maddikunta P, Gadekallu R, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial

images using real-time image processing and machine learning. *IET Image Processing*. (2021). doi: 10.1049/IPR2.12222
127. Abbas S, Jalil Z, Javed A, Batool I, Khan Z, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *J Comput Sci*. (2021) 7:e390. doi: 10.7717/peerj-cs.390
128. Deepa N, Prabadevi B, Maddikunta PK. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
129. Arkwright J, Blenman N, Underhill U. *In-vivo* demonstration of a high resolution optical fibre manometry catheter for diagnosis of gastrointestinal motility disorders. *Opt Expr*. (2009) 17:4500–8. doi: 10.1364/OE.17.004500
130. Mukred A, Singh D, Safie N. Examining the influence of perceived need on the adoption of information system in public hospitals in Yemen. *Asia-Pacific J Inform Technol Multimedia*. (2020) 9:35–49. doi: 10.17576/apjitm-2020-0902-04
131. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HM, et al. Hep-Pred: hepatitis C staging prediction using fine gaussian SVM. *Comput Mater Continua*. 69:2021. doi: 10.32604/cmc.2021.015436
132. Hasan MK, Islam S, Sulaiman R, Khan S, Hashim AH, Habib S, et al. Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access*. (2021) 9:47731–42. doi: 10.1109/ACCESS.2021.3061710
133. Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, et al. IoT for smart cities: machine learning approaches in smart healthcare—a review. *Future Internet*. (2021) 13:218. doi: 10.3390/fi13080218

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Upadhyaya, Hasan, Abdel-Khalek, Hassan, Srivastava, Sharan, Islam, Saad and Vo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

PERSPECTIVE
published: 03 December 2021
doi: 10.3389/fpubh.2021.756675

NextGen Public Health Surveillance and the Internet of Things (IoT)

Kirti Sundar Sahu[1], *Shannon E. Majowicz*[1], *Joel A. Dubin*[1,2] *and Plinio Pelegrini Morita*[1,3,4,5,6]*

[1] School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada, [2] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada, [3] Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada, [4] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada, [5] Ehealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada, [6] Research Institute for Aging, University of Waterloo, Waterloo, ON, Canada

Recent advances in technology have led to the rise of new-age data sources (e.g., Internet of Things (IoT), wearables, social media, and mobile health). IoT is becoming ubiquitous, and data generation is accelerating globally. Other health research domains have used IoT as a data source, but its potential has not been thoroughly explored and utilized systematically in public health surveillance. This article summarizes the existing literature on the use of IoT as a data source for surveillance. It presents the shortcomings of current data sources and how NextGen data sources, including the large-scale applications of IoT, can meet the needs of surveillance. The opportunities and challenges of using these modern data sources in public health surveillance are also explored. These IoT data ecosystems are being generated with minimal effort by the device users and benefit from high granularity, objectivity, and validity. Advances in computing are now bringing IoT-based surveillance into the realm of possibility. The potential advantages of IoT data include high-frequency, high volume, zero effort data collection methods, with a potential to have syndromic surveillance. In contrast, the critical challenges to mainstream this data source within surveillance systems are the huge volume and variety of data, fusing data from multiple devices to produce a unified result, and the lack of multidisciplinary professionals to understand the domain and analyze the domain data accordingly.

Keywords: real-time data, rapid surveillance, data source, big data, innovation

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Ebuka Ibeke,
Robert Gordon University, United Kingdom
Shweta Agrawal,
Sage University, India
Praveen Kumar,
VIT University, India

***Correspondence:**
Plinio Pelegrini Morita
plinio.morita@uwaterloo.ca

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *10 August 2021*
Accepted: *12 November 2021*
Published: *03 December 2021*

Citation:
Sahu KS, Majowicz SE, Dubin JA and Morita PP (2021) NextGen Public Health Surveillance and the Internet of Things (IoT). Front. Public Health 9:756675. doi: 10.3389/fpubh.2021.756675

INTRODUCTION

The function of public health systems is to understand and respond to health trends affecting populations (1). This is achieved through public health surveillance, that is, the ongoing collection and analysis of population health indicators. Traditional surveillance data collection can be cumbersome, expensive, and slow, often relying on paper-based and digitally extracted data sources. Social media and crowdsourcing are data sources that can be leveraged for surveillance data (2, 3). Sources like Twitter, Facebook, Google, and Reddit have been successfully used to explore behavior and health outcomes (4–6). These are now being accepted as potential data sources across several health domains (7, 8).

Another promising data source is the increasing number of devices (e.g., smart home monitors, wearables) and the technology to interconnect them. Internet of Things (IoT) technologies have become mainstream within communities and individual households (9). Wearables and sensors can track personalized parameters of healthy living, including sleep, physical activity, and sedentary behavior (10, 11). These devices can provide insights into population health, disease management, and active assisted living services (12, 13). IoT data has several advantages over traditional surveillance data: high volume and frequency of data collection, data triangulation, real-time availability, and minimal acquisition effort.

Existing literature discusses the potential use of the IoT data sources for different purposes within multiple domains including healthcare. Among healthcare domain, area specific application can be seen for pediatric, geriatrics, chronic disease supervision, private health, and fitness management (14, 15), but no single study exists to put together the views to utilize the IoT data with specific emphasis on public health surveillance.

This article summarizes the existing literature on the use of IoT as a data source for surveillance. We discuss the shortcomings of current data sources and how IoT can meet the needs of surveillance. Challenges facing the large-scale application of IoT data to surveillance are also explored.

PUBLIC HEALTH SURVEILLANCE AND CHALLENGES WITH EXISTING DATA SOURCES

Public health recommendations focus on the social determinants of health and health equity (16). Surveillance is the process by which ongoing health data are collected, analyzed, and reported, and it is critical to informing public health services. In 1968, the World Health Organization listed 10 essential data sources for surveillance (17) (**Figure 1**: Traditional data sources) that at the time relied on paper-based data collection and manual data entry. Surveillance capability has evolved enormously alongside advances in technology. It now includes digital data extracted from several sources (**Figure 1**: Modern data sources), offering reduced processing time, fewer errors, and reduced lag between data collection and its use.

The above said, surveillance data are still often obtained from questionnaire-based surveys online surveys, in-person or telephone-based interviews (18), and such data collection requires enormous resources and funding (19, 20). Data quality can be compromised by declining response rates (18), recall bias (21), and low granularity of the data (22) as in the traditional data collection system, there is a limited number of subjects provide their inputs. Without complete and comprehensive information, the value of the data reduced. For example, fewer subjects with a smaller *n*, really only impacts the precision of the estimates that come from surveillance. To further explain, the system might not get very precise incidence estimates, which may or may not be a problem depending on the goal of the system. The bigger issue with declining response rates is that they usually do not happen at random and meaning you're a less representative set of results. This is an issue if the factors that lead to making it into surveillance also relate to the issue you are trying to measure with the surveillance system. Current data used for the surveillance have challenges like missing data, under-reporting, inconsistencies, invalid data, illegible handwriting, non-standardization of vocabulary, measurement error, and inappropriate fields (23). Traditional data sources used in surveillance are often delayed. For example, at least 1 year is required for getting a Canadian Community Health Survey (CCHS) update. "Public Health Ontario" in Canada affirms interdependent gaps within surveillance, insufficient data to build comprehensive health indicators (24), and an absence of existing mechanisms to capture some of healthcare's vital components.

Current surveillance relies on both prospective and retrospective data collection, analysis, and reporting (25). The current pandemic has highlighted the essential need for real-time public health surveillance to improve the evidence-based decision-making process (26). Our evolving knowledge about chronic diseases, their risk factors, and management also demands the modernization of surveillance (25). Real-time responses to emerging public health threats require real-time and systematic data collection.

NEXT-GENERATION DATA SOURCES FOR PUBLIC HEALTH SURVEILLANCE

Researchers have attempted to build and analyze health indicators using innovative data sources (27–29). They are exploring the use of smartphones (30), online searches (31), social media (7), wearables (32), ambient sensors (33), electronic health records (EHRs) (27, 34), medical-administrative records (27), and pharmacy sales (28) to broaden the scope of surveillance.

As a source of surveillance data, information technologies are potentially advantageous because their near-universal uptake by a significant portion of the population creates vast quantities and varieties of data (22). For example, wearable data from six billion nights has been used to understand sleep duration, quality, and change in pattern with time (35, 36). Effective use of big data for surveillance requires innovative analytical methods such as data integration (32) and data visualization (28, 37, 38). Big data analytics is becoming mainstream in public health, integrating knowledge and skills from health informatics and biostatistics (39).

THE INTERNET OF THINGS AS A NOVEL DATA SOURCE

The Internet of Things (IoT) is a technological innovation through which devices can communicate with each other in real-time through an internet connection (40). For example, several household devices are interconnected to achieve a common objective, such as monitoring temperature or motion (40). Integrated devices can include different sensors, mobile phones, mobile applications,

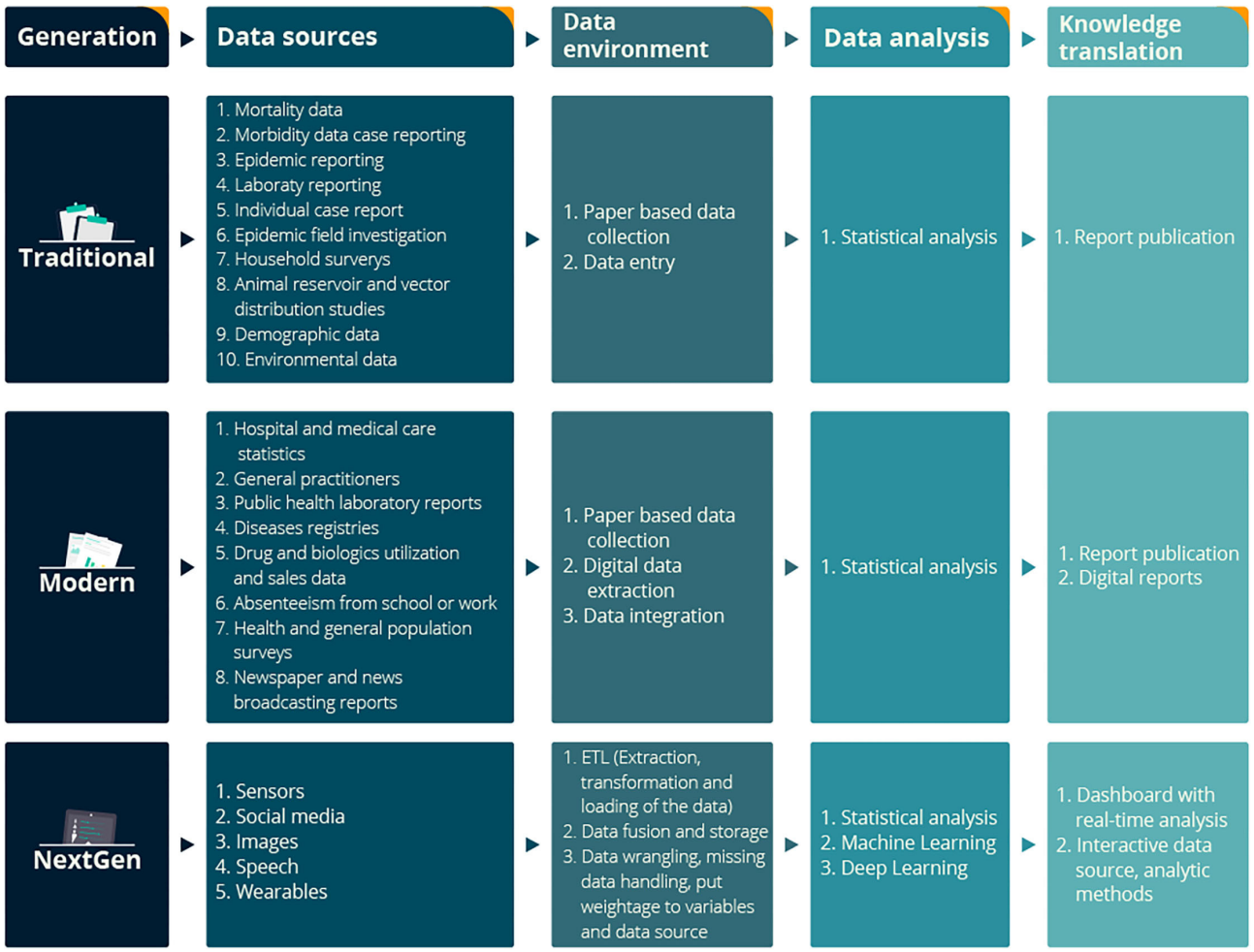


FIGURE 1 | Conceptual framework of NextGen Public Health Surveillance with Traditional, Modern, and NextGen data sources. Traditional and modern data sources extracted from Declich and Carter (17).

wearable devices, and Radio-Frequency Identification (RFID) tags (40).

IoT devices have accelerated data collection (13, 41). Connectivity among people, machines, and organizations increases as device availability and affordability improve (22). This increase in connectivity is because of the ease of use of the devices, user-friendly designs, and internet speed. These parameters reduced the time gap within communication, broaden the scope of communication by providing different choice, be it audio visual, text, or hybrid of multiple methods. People can interact with the machines and vice versa, which was not possible earlier due to lack of technological progress. In 2011, the number of interconnected devices overtook the actual number of people globally (42). The potential for data generation is exponential (41). As the IoT data has already been successfully used in multiple setups to monitor individual health outcomes and report on environmental conditions, some of the best use cases has been described below.

Use of IoT Data to Support Individual Health Outcomes

The management of chronic conditions has traditionally relied on patients interacting with their healthcare providers in person. However, patients spend most of their time outside the clinic. IoT monitoring provides an opportunity to collect real-time health information between patient-healthcare provider interactions.

Smart devices, such as wristbands, with IoT technology have been developed to measure individual physiological data, including physical activity (10, 43), sedentary time (44), oxygen saturation (45–47), heart rhythm (45, 46), muscle tremors (48), spinal posture (49), brainwaves (50), sleep (51), diet (52, 53), electrodermal activity monitoring for sympathetic response (44) and oral health care (54). With regards to specialized medical care, IoT technology has been used to cater to the need of cardiovascular (18), cardiopulmonary (18) and ophthalmology (55). With regards to different categories of populations, IoT has been used to help to monitor indicators related to women's health (56), including pregnancy (57), soldiers at the country borders

(58), nursing care at the hospitals (59), the elderly population in the long-term-care homes (60), persons with neurological conditions at the rehabilitation center (49), and also for persons with respiratory complaints including asthma (61).

IoT devices have a multipurpose use within the healthcare field, such as their capabilities can range from providing prenatal care to rehabilitation to monitoring seniors or athletes. IoT devices have successfully provided real-time health information on maternal and fetal health between regular appointments (57). By monitoring vital signs using sensors, IoT platforms have been designed to provide people with diabetes with feedback and notifications to mitigate the risk of complications (62–64). Additionally, wearable devices have been used to detect falls and changes in behavioral activity for seniors living independently (65–68). Monitoring systems have also been developed to evaluate sports rehabilitation (69–72). IoT can support individual outcomes by allowing patients to manage their health outside of the clinical setting.

Use of IoT Data to Monitor Environmental Conditions

The IoT can also monitor environmental conditions in areas where we live, work, and play. Monitoring air purification in hospital settings plays a role in mitigating hospital-related infections (73). Monitoring air quality is already used to quantify climate change impact (74) and has the potential to help mitigate its impact in the future (75). IoT has been employed to monitor hospital circulating air volume, ozone concentration, temperature, humidity, and leaked ultraviolet intensity (73). Preventive behavior like hand washing can also be monitored (76). Indicators of healthy outdoor environments, such as water pollution and air quality, have been another target of IoT health research (61, 77, 78).

THE INTERNET OF THINGS IN PUBLIC HEALTH SURVEILLANCE

IoT data has been successfully used in other health domains but has not yet been fully used in public health. In response to the pandemic, the 2020 Riyadh Declaration made several recommendations to address the shortcomings in global public health response systems (79). The Declaration prioritized the need for scalable and sustainable digital health technologies and the adoption of health intelligence (79). There is a growing interest in using IoT data for building public health indicators at various levels (80–82).

Advantages of IoT in Public Health Surveillance

IoT data have the potential to overcome shortcomings of current surveillance. IoT data sources provide high-frequency data with greater usability, and much of the device infrastructure for surveillance is already in place (i.e., smartphones, wearable technologies, internet access). Currently, worldwide more than three billion smartphone users (83), 722 million users of several kinds of wearable devices (84), and more than 1.2 billion smart-home connected devices exist (85). IoT data benefits from essential features like high granularity (22), objectivity (32), and validity (86). These "user-generated data ecosystems" are being generated with minimal effort by the device users and researchers. To date, the monetary cost to participants and researchers is low, suggesting that public health monitoring costs would likewise be minimal (87, 88). Finally, IoT enables near real-time data collection (89). This can significantly reduce the time gap between health events, data collection, reporting, and intervention.

Here we have assessed IoT's current attributes using the framework for evaluating public health surveillance by Groseclose et al. (90), which outlines nine features of surveillance systems to consider (**Table 1**). As summarized in the table, the major advantages of IoT data sources appear to be high-frequency data collection, the potential to have syndromic surveillance, zero effort data collection method, high volume, and variety of data. The major disadvantages appear to be lack of representativeness within a single data source, private players' involvement as the data owner, the need for a high technological system to store, clean, and analyze the data, and interoperability. In addition to the above points, data privacy concerns of users are a potential disadvantage of acceptance of this technology from the user point of view (81).

Challenges to Using IoT in Public Health Surveillance

The challenge now is how to access and analyze the data being gathered. Some IoT companies create sharable, research-oriented data sources, such as "donate your data" from ecobee, a smart thermostat company in Canada (91). ecobee's smart home products include motion and temperature sensors, and research teams have access to longitudinal data from thousands of households with a data granularity of 5 min intervals.

Other IoT companies publish studies from their own smart devices using artificial intelligence algorithms for population-level measurements. For example, Fitbit wearables recorded sleep data from over six billion nights of its customers' sleep (35), the most prominent sleep dataset ever collected. Similarly, Oura Health used IoT data gathered from their Oura ring, a wearable sensor that tracks key signals from the human body (sleep, heart rate, skin temperature, physical activity), delivering critical insights to help an individual harness their body's potential daily and also to monitor vital health indicators (92).

Another hurdle is the ability to fuse data from multiple devices to produce a unified result. Several research projects have focused on making IoT data fusion viable in the real world by designing computing infrastructure and data fusion techniques (89, 93). Real-time IoT analysis from multiple health monitoring devices may overwhelm current computational capabilities, such as using multiple devices to monitor each football player's physiological indicators during a game (94). A distributed computational framework to handle complex computational needs was developed by Higinio et al. for health surveillance (94).

TABLE 1 | Analysis of IoT as a data source for public health surveillance, using Groseclose et al. (90)[#] framework for evaluating public health surveillance.

Attributes (definition)	Features of IoT data
Simplicity *"The system's structure and ease of operation. The system should be as simple as possible."* (90)	• Data collection/extraction from users without complex interactions using Application Programming Interface (APIs) that the manufacturer often provides. • Easy access to the data, which is often collected by passive sensors, minimizing the burden for the user. • IoT systems rapidly generate large volumes of data in real-time, creating challenges associated with managing, hosting, and analyzing big data. • Diverse types of data being generated: numeric, images, text, or audio. • Collects vast amounts of data from the same individual, often supporting longitudinal analysis.
Flexibility *"Ability to adapt to changing information needs or technological operating conditions with little additional time, personnel, or allocated funds."* (90)	• Application Programming Interfaces (APIs) make it easy to adapt to the technology to the end-users being used, type of data, type of database, type of storage, and security requirements. • New IoT data sources that use APIs can easily be integrated into systems, also affording changes in a data structure as technologies evolve. • Changes in case definition can be updated in algorithms rather than requiring changes to data collected since systems can access the raw data. • The system can be automated to generate alert systems without manual effort which can help public health officials identify potential signals for future outbreaks early.
Data quality *"Completeness and validity of the data recorded in the system."* (90)	• IoT data often suffers from missing, inaccurate, and incomplete data. • Wearable sensors that require participants to recharge and remember to interact with the device often have larger volumes of missing data. • Ambient sensors often generate continuous and complete datasets as they are always connected, powered on, and streaming. • Technology development is leading to improved data quality across all IoT sensors.
Acceptability *"Willingness of persons and organizations to participate in the system."* (90)	• IoT technologies are pervasive, and in the community, a part of the population is already using those technologies to generate data. • IoT adoption is accelerating in the last decade and is predicted to be much higher in the near future. • Recent advancement in technology used "skin interfaced sensors" not only monitor physical activities and vital signs but also keep track of molecular biomarkers of the human body (104). • Users need to agree to share their data, as it has already been collected.
Sensitivity *"At the level of case reporting: the proportion of cases of a disease or event detected by the system. Ability to detect outbreaks over time and evaluation of surveillance system."* (90)	• IoT sensors, in most cases, do not focus on the detection of specific diseases such as COVID-19 or influenza but rather on symptoms like fever, abnormal heart rate, or change in gait pattern. • IoT technology is ideal for supporting syndromic surveillance by collecting data about healthy behaviors and health variables in real-time. • IoT technology will collect data that is often indirectly associated with health and health risk behaviors (e.g., indoor motion data to quantify sleep patterns, phone mobility data used to quantify response to COVID-10 policies). • IoT will provide extensive participant data with a higher likelihood of the presence of events. • The longitudinal nature of the data can detect future anomalies using Artificial Intelligence models within the healthcare sector and send alerts to policymakers. The longitudinal and continuity nature of the data will provide richer insights into population behaviors, which increases the likelihood of getting the events of interest.
Positive predictive value *"The proportion of reported cases that actually have the event under surveillance."* (90)	• The proportion of the presence of IoT within the community is increasing and predicting the true positive cases will be easier using IoT data by identifying early alerts. • The detection of specific diseases is possible, as technologies such as lab on a chip (106, 107) allow for real-time detection of pathogens and contaminants. • Positive predictive value seems to be in a disadvantageous position with the current IoT data environment, but this might change in the future.
Representativeness *"Ability to accurately describe the occurrence of a health-related event overtime and distribution of the population by place and person."* (90)	• Large number of participants can provide access to data who were not represented in the traditional data collection method. • IoT technologies are ubiquitous, highly pervasive, and are generating data 24/7. • Data mining from sensors already owned by the population generates a biased sample, with data from the wealthier and more physically active part of the population. • Studies can supplement biased samples by deploying targeted studies to collect data from under-represented subgroups of the population.
Timeliness *"Reflects the speed between steps in a system."* (90)	• Data is often collected at high frequencies, often affording access to data in the near real-time. • An increase in the data's granularity and the longitudinal nature of the data can provide richer insights, for instance, faster alerts of anomalies for specific health issues, and support the creation of innovative indicators. • In the near future, the IoT data source may become helpful to identify future pandemic and climate-related emergencies. Immediate assessment of the impact of policy changes (for example, "work from home" during the pandemic) can be possible using IoT data. • Improvement from traditional data sources where data collection often happens once yearly or less frequently.
Stability *"Ability to rely on the system for availability and to collect manage and provide data without failure. Ability to be operational when needed."* (90)	• Private cloud systems can provide the necessary data security and maintain the users' privacy. • Redundant, always available, more stable public health surveillance platforms/systems can be built using private cloud solutions, having the capacity to collect uninterrupted data without failure. IoT manufacturers and IoT data custodians can deliver such redundant and stable systems for their consumers' everyday use. • The disadvantage of these IoT data manufacturers is ever-changing company environment (for example, corporate and big private entities) might not provide a stable source of data. The alternative source of data should be listed as a backup plan to support and strengthen when required.

[#] *Groseclose et al. (90).*

The use of each smart devices' computing capabilities effectively shared advanced health monitoring applications (94).

Regarding technical challenges related to IoT, some of the critical issues are energy optimization, hardware compatibility, security, and data connectivity (95). A recent study by Iwendi et al. in 2020 shows that there are certain highly specialized algorithm such as a "hybrid meta-heuristic algorithm" has the potential to optimize the energy consumption of the sensors related to wireless sensor networks (95).

Aberration detection identifies unusual incidents or information trends with possible significance to clinical or public health (96). Methods for detecting such aberrations have also evolved significantly. Current modeling methods can now analyze individual surveillance data collected from different sources and integrate multiple covariates (97). The algorithms used for signal recognition have improved over the last decade and are now better equipped to utilize advanced informatics to capture surveillance data aberrations (96, 97) accurately.

In 2018, Faverjon C. and Berezowski J. elaborated on IoT data's utility for aberration detection (97, 98). Two studies have shown that user data from wearables (Fitbit and the Oura ring) could detect early signs of COVID-19 infection (99–103). Evidence shows the risk of hospitalization related to COVID-19 can be calculated from self-reported symptoms and predictive physiological signs by combining different health and behavioral data from consumer wearable devices; this may help identify pathological changes weeks before observation using traditional epidemiological monitoring (99, 100). As described in the study using Fitbit wearable, it has the potential to detect almost half of COVID-19 positive cases 24 h before participants reported the onset of symptoms with 70 percent specificity (103). Besides joint effort by multiple countries to develop vaccines and potential drugs to prevent and treat COVID-19, skin-integrated and skin interfaced sensors, positioned at optimal locations of the body, might address the ongoing and critical need for objective, continuous, and sensitive tools to detect COVID-19 symptoms early in the general population (101, 104). A research study highlighted a practical approach for managing epidemics using digital technologies with a roadmap to a rapid and universal diagnostic method for the population level detection of several respiratory infections in advance of symptoms (102). These anomalies could predict future outbreaks (97) and prevent the spread of infectious diseases (105).

NEXTGEN PUBLIC HEALTH SURVEILLANCE

The COVID-19 pandemic has revealed a need to strengthen our public health surveillance and response systems. With the availability of public data and advances in collection and analysis, there is an opportunity to strengthen existing surveillance systems by harnessing complementary data sources like IoT-based data (31).

Figure 1 describes the NextGen surveillance systems' conceptual framework. The first layer describes the sources of public health data. The second layer represents the data architecture. Once the data integration process is completed, data manipulation and analysis can be possible using statistics, machine learning, and deep learning algorithms. This process will help discover new public health indicators and advance our understanding of existing disease risk factors.

CONCLUSION

Current public health surveillance systems have unique challenges in getting the relevant data at the right time and utilizing those data sources for policy-level decision-making. There is a considerable volume of non-traditional data being self-generated by the public through their ubiquitous use of smart devices. Public health has the potential to utilize the real-time, longitudinal data collected through the Internet of Things (IoT) necessary for health surveillance. Advances in computing are now bringing IoT-based surveillance into the realm of possibility. The advantages of IoT data include high-frequency, high volume, zero effort data collection method, with a potential to have syndromic surveillance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

KS and PM developed the theoretical framework for the paper. PM supervised KS for the project. KS wrote the first draft of the manuscript with input from all authors. SM and JD helped provide overall direction and planning. All authors contributed to manuscript reading and revision and have approved the submitted version.

FUNDING

This work was supported by Natural Sciences and Engineering Research Council of Canada (RPGIN-2017-05310) and the Ontario Centre of Innovation (33080).

ACKNOWLEDGMENTS

Thanks to all the team members of Ubilab, at the University of Waterloo, for their support.

REFERENCES

1. Choi BCK. The past, present, and future of public health surveillance. *Scientifica.* (2012) 2012:1–26. doi: 10.6064/2012/875253
2. Wazny K. Applications of crowdsourcing in health: an overview. *J Glob Health.* (2018) 8:010502. doi: 10.7189/jogh.08.010502
3. Kraemer JD, Strasser AA, Lindblom EN, Niaura RS, Mays D. Crowdsourced data collection for public health: a comparison with

nationally representative, population tobacco use data. *Prev Med.* (2017) 102:93–9. doi: 10.1016/j.ypmed.2017.07.006
4. Dalton CB. Enablers of innovation in digital public health surveillance: lessons from flutracking. *Int Health.* (2017) 9:145–7. doi: 10.1093/inthealth/ihx009
5. Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: facebook likes. *J Med Internet Res.* (2015) 17:e98. doi: 10.2196/jmir.3970
6. Bagroy S, Kumaraguru P, De Choudhury M. *A Social Media Based Index of Mental Well-Being in College Campuses*. New York, NY: ACM Press (2017).
7. Edo-Osagie O, De La Iglesia B, Lake I, Edeghere O. A scoping review of the use of twitter for public health research. *Comput Biol Med.* (2020) 122:103770. doi: 10.1016/j.compbiomed.2020.103770
8. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. *Am J Public Health.* (2017) 107:e1–8. doi: 10.2105/AJPH.2016.303512a
9. Plageras AP, Psannis KE, Ishibashi Y, Kim BG. IoT-based surveillance system for ubiquitous healthcare. In: *IECON Proceedings (Industrial Electronics Conference)* (Florence: IEEE Computer Society).
10. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y. Examining sensor-based physical activity recognition and monitoring for healthcare using internet of things: a systematic review. *J Biomed Inform.* (2018) 87:138–53. doi: 10.1016/j.jbi.2018.09.002
11. Sahu KS, Oetomo A, Morita PP. Enabling remote patient monitoring through the use of smart thermostat data in canada: exploratory study. *JMIR Mhealth Uhealth.* (2020) 8:e21016. doi: 10.2196/21016
12. Talal M, Zaidan AA, Zaidan BB, Albahri AS, Alamoodi AH, Albahri OS, et al. Smart home-based iot for real-time and secure remote health monitoring of triage and priority system using body sensors: multi-driven systematic review. *J Med Syst.* (2019) 43:42. doi: 10.1007/s10916-019-1158-z
13. Asghari P, Rahmani AM, Javadi HHS. Internet of things applications: a systematic review. *Comput Networks.* (2019) 148:241–61. doi: 10.1016/j.comnet.2018.12.008
14. Bhuiyan MN, Rahman MM, Billah MM, Saha D. Internet of things (IoT): a review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities. *IEEE Internet Things J.* (2021) 8:10474–98. doi: 10.1109/JIOT.2021.3062630
15. Riazul Islam SM, Kwak D, Humaun Kabir M, Hossain M, Kwak KS. The internet of things for health care: a comprehensive survey. *IEEE Access.* (2015) 3:678–708. doi: 10.1109/ACCESS.2015.2437951
16. DeSalvo KB, O'Carroll PW, Koo D, Auerbach JM, Monroe JA. Public health 3.0: time for an upgrade. *Am J Public Health.* (2016) 106:621–2. doi: 10.2105/AJPH.2016.303063
17. Declich S, Carter AO. Public health surveillance: historical origins, methods and evaluation. *Bull World Health Organ.* Geneva (1994) 72:285–304.
18. IOM (Institute of Medicine). *A Nationwide Framework for Surveillance of Cardiovascular and Chronic Lung Diseases*. Washington, DC: National Academies Press (2011).
19. Frieden TR, Harold Jaffe DW, Thacker SB, Moolenaar RL, Lee LM, Meyer PA, et al. *CDC's Vision for Public Health Surveillance in the 21st Century*. Atlanta: Centers for Disease Control and Prevention(2012).
20. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. *Int J Behav Healthc Res.* (2011) 2:320–32. doi: 10.1504/IJBHR.2011.043414
21. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol.* (1990) 43:87–91. doi: 10.1016/0895-4356(90)90060-3
22. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis.* (2016) 214:S375–9. doi: 10.1093/infdis/jiw400
23. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. *Int J Environ Res Public Health.* (2014) 11:5170–207. doi: 10.3390/ijerph110505170
24. Public Health Ontario. *Gaps in Public Health Indicators and Data in Ontario*. Revised ed. Toronto, ON (2016). Available online at: https://www.publichealthontario.ca/-/media/documents/public-health-indicator-gaps.pdf?la=en (accessed February 22, 2019).
25. Chretien JP, Tomich NE, Gaydos JC, Kelley PW. Real-time public health surveillance for emergency preparedness. *Am J Public Health.* (2009) 99:1360–3. doi: 10.2105/AJPH.2008.133926
26. Buehler JW, Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century. Introduction. *MMWR Suppl.* (2012) 61:1–2.
27. Seeskin ZH, LeClere F, Ahn J, Williams J. *Uses of Alternative Data Sources for Public Health Statistics and Policymaking: Challenges and Opportunities*. (2018). Available online at: https://www.norc.org/PDFs/Publications/SeeskinZ_UsesofAlternativeDataSources_2018.pdf (accessed August 24, 2020).
28. Groseclose SL, Buckeridge DL. Public health surveillance systems: recent advances in their use and evaluation. *Annu Rev Public Health.* (2017) 38:57–79. doi: 10.1146/annurev-publhealth-031816-044348
29. Althouse BM, Scarpino S V, Meyers LA, Ayers JW, Bargsten M, Baumbach J, et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Sci.* (2015) 4:17. doi: 10.1140/epjds/s13688-015-0054-0
30. Navin K, Krishnan MBM, Lavanya S, Shanthini A. A mobile health based smart hybrid epidemic surveillance system to support epidemic control programme in public health informatics. In: *IEEE International Conference on IoT and its Applications, ICIOT*. (Nagapattinam: Institute of Electrical and Electronics Engineers Inc.) (2017).
31. Mancuso M, Yao X, Otchere D, Rasali D, Clark E, Svenson LW, et al. Proof of concept paper: non-traditional data sources for public health surveillance. In: *DH - Proceedings of the 2016 Digital Health Conference.* New York, NY: Association for Computing Machinery, Inc. (2016).
32. LaMunion SR, Fitzhugh EC, Crouter SE. Challenges and opportunities related to the objective assessment of physical activity within U.S. health surveys. *Ann Epidemiol.* (2020) 43:1–10. doi: 10.1016/j.annepidem.2020.01.011
33. Lamiae E, Fatiha E, Mohammed B, Hicham GT. *Contribution to a Smart Home Design for Medical Surveillance.* Cham: Springer.
34. Huang J, Loschen W. Potential applications of emerging technologies in disease surveillance. *Online J Public Health Inform.* (2019) 11:e340. doi: 10.5210/ojphi.v11i1.9821
35. Pogue D. *What Fitbit Discovered From 6 Billion Nights of Sleep Data. Yahoo*. (2018). Available online at: https://uk.finance.yahoo.com/news/exclusive-fitbits-6-billion-nights-sleep-data-reveals-us-110058417.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cud2Vmb3J1bS (accessed March 2, 2020).
36. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res.* (2019) 21:e16273. doi: 10.2196/16273
37. DeSalvo KB, Wang YC, Harris A, Auerbach J, Koo D, O'Carroll P. Public health 3.0: a call to action for public health to meet the challenges of the 21st century. *Prev Chronic Dis.* (2017) 14:170017. doi: 10.5888/pcd14.170017
38. Zeng D, Cao Z, Neill DB. Artificial intelligence–enabled public health surveillance—from local detection to global epidemic monitoring and control. *Artif Intell Med.* (2021):437–53. doi: 10.1016/B978-0-12-821259-2.00022-3
39. Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. *J Epidemiol Community Health.* (2020) 74:612–6. doi: 10.1136/jech-2018-211654
40. Atzori L, Iera A, Morabito G. The internet of things: a survey. *Comput Networks.* (2010) 54:2787–805. doi: 10.1016/j.comnet.2010.05.010
41. Steele R, Clarke A. The internet of things and next-generation public health information systems. *Commun Netw.* (2013) 5:4–9. doi: 10.4236/cn.2013.53B1002
42. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of things (IoT): a vision, architectural elements, and future directions. *Futur Gener Comput Syst.* (2013) 29:1645–60. doi: 10.1016/j.future.2013.01.010
43. Subasi A, Radhwan M, Kurdi R, Khateeb K. IoT based mobile healthcare system for human activity recognition. In: *15th Learning and Technology Conference, L and T*. Jeddah: Institute of Electrical and Electronics Engineers Inc. (2018).
44. Hilty DM, Armstrong CM, Luxton DD, Gentry MT, Krupinski EA. A scoping review of sensors, wearables, and remote monitoring for behavioral health: uses, outcomes, clinical competencies, and research directions. *J Technol Behav Sci.* (2021) 6:1–36. doi: 10.1007/s41347-021-00199-2

45. Marques G, Roque Ferreira C, Pitarma R. A system based on the internet of things for real-time particle monitoring in buildings. *Int J Environ Res Public Health.* (2018) 15:821. doi: 10.3390/ijerph15040821
46. Li SH, Lin BS, Wang CA, Yang CT, Lin BS. Design of wearable and wireless multi-parameter monitoring system for evaluating cardiopulmonary function. *Med Eng Phys.* (2017) 47:144–150. doi: 10.1016/j.medengphy.2017.06.009
47. Adiputra RR, Hadiyoso S, Hariyani YS. Internet of things: low cost and wearable SpO2 device for health monitoring. *Int J Electr Comput Eng.* (2018) 8:939–45. doi: 10.11591/ijece.v8i2.pp939-945
48. Chiuchisan I, Geman O. An approach of a decision support and home monitoring system for patients with neurological disorders using internet of things concepts. *WSEAS Trans Syst.* (2014) 13:460–9.
49. Nam KH, Kim DH, Choi BK, Han IH. Internet of things, digital biomarker, and artificial intelligence in spine: current and future perspectives. *Neurospine.* (2019) 16:705–11. doi: 10.14245/ns.1938388.194
50. de Oliveira WG Jr, de Oliveira JM, Munoz R, de Albuquerque VHC. A proposal for internet of smart home things based on BCI system to aid patients with amyotrophic lateral sclerosis. *Neural Comput Appl.* (2020) 32:11007–17. doi: 10.1007/s00521-018-3820-7
51. Surantha N, Kusuma GP, Isa SM. Internet of things for sleep quality monitoring system: a survey. In: *Proceedings - 11th. International Conference on Knowledge, Information and Creativity Support Systems, KICSS*. Yogyakarta: Institute of Electrical and Electronics Engineers Inc. (2016).
52. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access.* (2020) 8:28462–74. doi: 10.1109/ACCESS.2020.2968537
53. Mantey EA, Zhou C, Anajemba JH, Okpalaoguchi IM, Chiadika OD. Blockchain-secured recommender system for special need patients using deep learning. *Front Public Heal.* (2021) 9:737269. doi: 10.3389/fpubh.2021.737269
54. Salagare S, Prasad R. An overview of internet of dental things: new frontier in advanced dentistry. *Wirel Pers Commun.* (2020) 110:1345–71. doi: 10.1007/s11277-019-06790-4
55. Ting DSW, Lin H, Ruamviboonsuk P, Wong TY, Sim DA. Artificial intelligence, the internet of things, and virtual clinics: ophthalmology at the digital translation forefront. *Lancet Digit Heal.* (2020) 2:e8–9. doi: 10.1016/S2589-7500(19)30217-1
56. Yoldemir T. Internet of things and women's health. *Climacteric.* (2020) 23:423–5. doi: 10.1080/13697137.2020.1811563
57. Saarikko J, Niela-Vilen H, Ekholm E, Hamari L, Azimi I, Liljeberg P, et al. Continuous 7-month internet of things-based monitoring of health parameters of pregnant and postpartum women: prospective observational feasibility study. *JMIR Form Res.* (2020) 4:e12417. doi: 10.2196/12417
58. Patil N, Iyer B. Health monitoring and tracking system for soldiers using Internet of Things(IoT). In: *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA*. Greater Noida: Institute of Electrical and Electronics Engineers Inc. (2017).
59. Mieronkoski R, Azimi I, Rahmani AM, Aantaa R, Terävä V, Liljeberg P, et al. The Internet of Things for basic nursing care—A scoping review. *Int J Nurs Stud.* (2017) 69:78–90. doi: 10.1016/j.ijnurstu.2017.01.009
60. Finch M, Griffin K, Pacala JT. Reduced healthcare use and apparent savings with passive home monitoring technology: a pilot study. *J Am Geriatr Soc.* (2017) 65:1301–05. doi: 10.1111/jgs.14892
61. Isaac N, Sampath N, Gay V. SAM smart asthma monitoring: focus on air quality data and Internet of Things (IoT). In: *International Symposium on Medical Information and Communication Technology, ISMICT*. Sydney: IEEE Computer Society.
62. Alelyani S, Ibrahim A. Internet-of-things in telemedicine for diabetes management. In: *15th Learning and Technology Conference, L and T*. Institute of Electrical and Electronics Engineers Inc.) (2018).
63. Bruen D, Delaney C, Florea L, Diamond D. Glucose sensing for diabetes monitoring: recent developments. *Sensors.* (2017) 17:1866. doi: 10.3390/s17081866
64. Al-Taee MA, Al-Nuaimy W, Al-Ataby A, Muhsin ZJ, Abood SN. Mobile health platform for diabetes management based on the internet-of-things. In: *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT*. The Dead Sea: Institute of Electrical and Electronics Engineers Inc. (2015).
65. Ben Hassen H, Dghais W, Hamdi B. An E-health system for monitoring elderly health based on internet of things and fog computing. *Heal Inf Sci Syst.* (2019) 7:24. doi: 10.1007/s13755-019-0087-z
66. Hosseinzadeh M, Koohpayehzadeh J, Ghafour MY, Ahmed AM, Asghari P, Souri A, et al. An elderly health monitoring system based on biological and behavioral indicators in internet of things. *J Ambient Intell Humaniz Comput.* (2020) 1:3. doi: 10.1007/s12652-020-02579-7
67. Yacchirema D, De Puga JS, Palau C, Esteve M. Fall detection system for elderly people using IoT and big data. *Procedia Comput Sci*. (2018) 130:603–10. doi: 10.1016/j.procs.2018.04.110
68. Dohr A, Modre-Opsrian R, Drobics M, Hayn D, Schreier G. The internet of things for ambient assisted living. In: S*eventh International Conference on Information Technology:* New Generations (Porto: IEEE) (2010).
69. Zhan K. Sports and health big data system based on 5G network and Internet of Things system. *Microprocess Microsyst.* (2021) 80:103363. doi: 10.1016/j.micpro.2020.103363
70. Zhao Y, You Y. Design and data analysis of wearable sports posture measurement system based on Internet of Things. *Alexandria Eng J.* (2021) 60:691–701. doi: 10.1016/j.aej.2020.10.001
71. Liu L. Construction of youth public sports service system based on embedded system and wireless IoT. *Microprocess Microsyst.* (2021) 83:103984. doi: 10.1016/j.micpro.2021.103984
72. Jiang Y. Combination of wearable sensors and internet of things and its application in sports rehabilitation. *Comput Commun.* (2020) 150:167–76. doi: 10.1016/j.comcom.2019.11.021
73. Yang L, Yao T, Liu G, Sun L, Yang N, Zhang H, et al. Monitoring and control of medical air disinfection parameters of nosocomial infection system based on Internet of Things. *J Med Syst.* (2019) 43:126. doi: 10.1007/s10916-019-1205-9
74. USGCRP. Impacts, risks, adaptation in the United States: the fourth national climate assessment. Volume II. In: Reidmiller DR, Avery CW, Easterling DR, Kunkel KE, Lewis KLM, Maycock TK, Stewart B. *U.S. Global Change Research Program*. Washington, DC: United Nations Office for Disaster Risk Reduction (2018).
75. Múnera D, Diana V, Aguirre J, Gómez NG. IoT-based air quality monitoring systems for smart cities: a systematic mapping study. *Int J Electr Comput Eng.* (2021) 11:3470–82. doi: 10.11591/ijece.v11i4.pp3470-3482
76. Bal M, Abrishambaf R. A system for monitoring hand hygiene compliance based-on Internet-of-Things. In: *Proceedings of the IEEE International Conference on Industrial Technology*. Toronto: Institute of Electrical and Electronics Engineers Inc.
77. Shafi U, Mumtaz R, Anwar H, Qamar AM, Khurshid H. Surface water pollution detection using Internet of Things. In: *15th International Conference on Smart Cities: Improving Quality of Life Using ICT and IoT, HONET-ICT*. Institute of Electrical and Electronics Engineers Inc. (2018).
78. Agarwal A, Shukla V, Singh R, Gehlot A, Garg V. Design and development of air and water pollution quality monitoring using iot and quadcopter. *Adv Intell Syst Comput*. (2018) 624:485–92. doi: 10.1007/978-981-10-5903-2_49
79. Al Knawy B, Adil M, Crooks G, Rhee K, Bates D, Jokhdar H, et al. The riyadh declaration: the role of digital health in fighting pandemics. *Lancet.* (2020) 396:1537–9. doi: 10.1016/S0140-6736(20)31978-4
80. Aiello AE, Renson A, Zivich PN. Social media-and internet-based disease surveillance for public health. *Annu Rev Public Heal.* (2020) 41:101–18. doi: 10.1146/annurev-publhealth-040119-094402
81. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. *Annu Rev Public Health.* (2018) 39:95–112. doi: 10.1146/annurev-publhealth-040617-014208
82. Hung K, Zhang YT. Implementation of a WAP-based telemedicine system for patient monitoring. *IEEE Trans Inf Technol Biomed.* (2003) 7:101–7. doi: 10.1109/TITB.2003.811870
83. Statista. *Smartphone Users*. (2020). Available online at: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ (accessed April 16, 2021).

84. Statista. *Global Connected Wearable Devices 2016–2022*. (2021). Available online at: https://www.statista.com/statistics/487291/global-connected-wearable-devices/ (accessed April 16, 2021).
85. Statista. *Global Smart Home Market Size 2016–2022*. (2021). Available online at: https://www.statista.com/statistics/682204/global-smart-home-market-size/ (accessed April 16, 2021).
86. Park HA, Jung H, On J, Park SK, Kang H. Digital epidemiology: use of digital data collected for non-epidemiological purposes in epidemiological studies. *Healthc Inform Res.* (2018) 24:253. doi: 10.4258/hir.2018.24.4.253
87. Grace SL, Taherzadeh G, Jae Chang IS, Boger J, Arcelus A, Mak S, et al. Perceptions of seniors with heart failure regarding autonomous zero-effort monitoring of physiological parameters in the smart-home environment. *Hear Lung J Acute Crit Care.* (2017) 46:313–9. doi: 10.1016/j.hrtlng.2017.04.007
88. Mihailidis A, Boger J, Hoey J, Jiancaro T. Zero effort technologies: considerations, challenges, and use in health, wellness, and rehabilitation. *Synth Lect Assist Rehabil Heal Technol.* (2011) 1:1–94. doi: 10.2200/S00380ED1V01Y201108ARH002
89. Akbar A, Kousiouris G, Pervaiz H, Sancho J, Ta-Shma P, Carrez F, et al. Real-time probabilistic data fusion for large-scale IoT applications. *IEEE Access.* (2018) 6:10015–27. doi: 10.1109/ACCESS.2018.2804623
90. Groseclose SL, German RR, Nsubuga P. Evaluating Public Health Surveillance. *Prin Pract Publ Health Surveil.* (2010) 166–197. doi: 10.1093/acprof:oso/9780195372922.003.0008
91. ecobee. *Donate your Data | ecobee | Smart Home Technology*. (2018). Available online at: https://www.ecobee.com/donateyourdata/ (accessed January 24, 2019).
92. Koskimäki H, Kinnunen H, Kurppa T, Röning J. *How Do We Sleep - a Case Study of Sleep Duration and Quality Using Data from Oura Ring*. New York, NY: Association for Computing Machinery, Inc. (2018).
93. Baloch Z, Shaikh FK, Unar MA. A context-aware data fusion approach for health-IoT. *Int J Inf Technol.* (2018) 10:241–5. doi: 10.1007/s41870-018-0116-1
94. Mora H, Gil D, Terol RM, Azorín J, Szymanski J. An IoT-Based computational framework for healthcare monitoring in mobile environments. *Sensors.* (2017) 17:2302. doi: 10.3390/s17102302
95. Iwendi C, Kumar Reddy PM, Reddy TG, Lakshmanna K, Kashif Bashir A, Piran J. A Metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw Pract Exp*. (2020) 51: 2558–71. doi: 10.1002/spe.2797
96. Yuan M, Boston-Fisher N, Luo Y, Verma A, Buckeridge DL. A systematic review of aberration detection algorithms used in public health surveillance. *J Biomed Inform.* (2019) 94:103181: doi: 10.1016/j.jbi.2019.103181
97. Faverjon C, Berezowski J. Choosing the best algorithm for event detection based on the intended application: a conceptual framework for syndromic surveillance. *J Biomed Inform.* (2018) 85:126–35. doi: 10.1016/j.jbi.2018.08.001
98. Jalali N, Sahu K, Oetomo A, Morita P. Understanding user behavior through the use of unsupervised anomaly detection: proof of concept using internet of things smart home thermostat data for improving public health surveillance. *JMIR Mhealth Uhealth.* (2020) 8:e21209: doi: 10.2196/21209
99. Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *npj Digit Med.* (2020) 3:156. doi: 10.1038/s41746-020-00363-7
100. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real-time. *Sci Adv*. (2021) 7:eabd6989. doi: 10.1126/sciadv.abd6989
101. Jeong H, Rogers JA, Xu S. Continuous on-body sensing for the COVID-19 pandemic: gaps and opportunities. *Sci Adv*. (2020) 6:eabd4794. doi: 10.1126/sciadv.abd4794
102. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Early detection of COVID-19 using a smartwatch. *medRxiv*. (2020) eabd4794. doi: 10.1101/2020.07.06.20147512
103. Fitbit. *Early Findings From Fitbit COVID-19 Study Suggest Fitbit Devices Can Identify Signs of Disease at Its Earliest Stages - Fitbit Blog*. (2020). Available online at: https://blog.fitbit.com/early-findings-covid-19-study/ (accessed September 10, 2020).
104. Xu C, Yang Y, Gao W. Skin-Interfaced Sensors in Digital Medicine: from Materials to Applications. *Matter*. (2020) 2:1414–45. doi: 10.1016/j.matt.2020.03.020
105. Rahman MS, Peeri NC, Shrestha N, Zaki R, Haque U, Hamid SHA. Defending against the novel coronavirus (COVID-19) outbreak: how can the Internet of Things (IoT) help to save the world? *Heal Policy Technol.* (2020) 9:136–8. doi: 10.1016/j.hlpt.2020.04.005
106. Nikoleli GP, Siontorou CG, Nikolelis DP, Bratakou S, Karapetis S, Tzamtzis N. Biosensors based on microfluidic devices lab-on-a-chip and microfluidic technology. *Nanotechnol Biosensors.* (2018) 375–94. doi: 10.1016/B978-0-12-813855-7.00013-1
107. Patra B, Sharma M, Hale W, Utz M. Time-resolved non-invasive metabolomic monitoring of a single cancer spheroid by microfluidic NMR. *Sci Rep.* (2021) 11:53. doi: 10.1038/s41598-020-79693-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Sahu, Majowicz, Dubin and Morita. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fpubh.2021.788347

Trustworthy Intrusion Detection in E-Healthcare Systems

Faiza Akram[1], *Dongsheng Liu*[1], *Peibiao Zhao*[1], *Natalia Kryvinska*[2]*, *Sidra Abbas*[3]* and *Muhammad Rizwan*[4]

[1] Department of Mathematics, School of Science, Nanjing University of Science and Technology, Nanjing, China, [2] Department of Information Systems, Faculty of Management, Comenius University in Bratislava, Bratislava, Slovakia, [3] Department of Computer Science, COMSATS University, Islamabad, Pakistan, [4] Department of Computer Science, Kinnaird College for Women, Lahore, Pakistan

In Internet of Things (IoT)-based network systems (IoT-net), intrusion detection systems (IDS) play a significant role to maintain patient health records (PHR) in e-healthcare. IoT-net is a massive technology with security threats on the network layer, as it is considered the most common source for communication and data storage platforms. The security of data servers in all sectors (mainly healthcare) has become one of the most crucial challenges for researchers. This paper proposes an approach for effective intrusion detection in the e-healthcare environment to maintain PHR in a safe IoT-net using an adaptive neuro-fuzzy inference system (ANFIS). In the proposed security model, the experiments present a security tool that helps to detect malicious network traffic. The practical implementation of the ANFIS model on the MATLAB framework with testing and training results compares the accuracy rate from the previous research in security.

Keywords: network security, privacy, ANFIS, intrusion detection, IoT based networks

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Randhir Kumar, Bennett University, India
Misbah Abbas, University of Science and Technology of China, China
Muhammad Baqer Mollah, Nanyang Technological University, Singapore
Praveen Kumar, VIT University, India

***Correspondence:**
Natalia Kryvinska
natalia.kryvinska@fm.uniba.sk
Sidra Abbas
sidra.abbas708@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *02 October 2021*
Accepted: *25 October 2021*
Published: *03 December 2021*

Citation:
Akram F, Liu D, Zhao P, Kryvinska N, Abbas S and Rizwan M (2021) Trustworthy Intrusion Detection in E-Healthcare Systems. Front. Public Health 9:788347. doi: 10.3389/fpubh.2021.788347

INTRODUCTION

Internet of Things (IoT)-based network systems (IoT-net) are considered as emerging advancements in the field of technology, where cloud network-based servers provide communication, storage, and problem-solving facilities, but these sorts of systems also contain security threats and issues as well (1–4). The network user can access its facility by using an internet source (5). The multiple hardware and software-based environments provide data and information to its end users. Some of the most prominent organizations like Apple, HP, Amazon, IBM, Oracle, Intel, etc., use cloud computing (CC) techniques. CC is based on three-layer models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Similarly, cloud networks are based on four types: private, public, hybrid, and community (6, 7).

In a cloud networking environment, problems in security and authorization are the key risks. Likewise, there are numerous risks which users and network service providers both face. Mainly, security issues arise from the data storage and networking side (8, 9). Many security enhancement-based algorithms have been proposed to make the IoT-net servers secure. Many cryptographic algorithms like RSA, AES, CRT-RSA, DES, blockchain, machine learning, and artificial intelligence-based code have been proposed to enhance security. Homomorphic encryption algorithms (10–12) help to provide better security to detect non-authorization factors. Many machine learning (13) and artificial intelligence-based security algorithms help to provide better data networking with new classification and risk assessment techniques.

However, many elements remain unsolved and require more improvement and advancement in areas such as cost management, resource utilization, speed prediction, and security, most importantly.

IoT also plays a magnificent role in the healthcare sector concerning data storage, online systems, software, laboratories, pathology, clinic sessions, etc. Medical patient health records (PHR) are an emerging version of IT and smart healthcare records (5). However, here we encounter problems of data security and privacy of PHR (1, 14–18). To overcome the PHR security issue, many researchers have been working on using blockchain and many other artificial intelligence techniques, which allows clinical expertise, patients, laboratories, and the world to be connected.

One of the most effective artificial intelligence techniques is called the artificial neural inference system (ANFIS), which is the combination of an artificial neural network (ANN) and a fuzzy inference system (FIS). ANFIS is mainly used as a computational model for resolving uncertainties, reasoning, and reducing security threats from networks and cloud servers (1). It helps the systems and servers examine the risk of data and information. ANNs work on the principle of mathematical calculation. The ANFIS model is based on if-then rule statements and crisp values, and the final ANFIS surface model uses result accuracy rate. It uses most of its computational time for data classification and estimation.

Problem Statement

In smart e-healthcare systems, networking servers provide a better intermediary platform for data storage, communication, and many other aspects. The end users (doctors, clinical experts, patients, laboratories) get the opportunity to access the PHR and access the cloud servers. It can be any authorized person. It is essential to detect and classify malicious activities and network traffic. The classification of network traffic and attack type detection helps make the system more secure and detect network intrusion.

Motivation

In this research paper, we discuss the existing security risks in IoT-net-based PHR. We classify these issues in managing risk, end-user risk, organizational risk, privacy risk, and many more. Our main agenda is to propose a less risky cloud network for everyone. We represent the list of shortcomings in existing malicious network traffic, and this will benefit the networking service provider and security handler community to comprehend the security issue.

Contribution

In this research, we try to implement the ANFIS model to detect network attacks on database servers. The main agenda is to detect the unauthorized access of users by using an ANFIS-based intrusion detection system. The proposed security algorithm helps to collect malicious activities or information from network traffic. Based on the if-then rule statement and ANFIS-based data classification which can help detect the intrusion attack, the system can determine if the cloud server has been hit by an intruder or not. If there is an attack, identifying the type of attack helps researchers implement precautionary measures to overcome the loss or block the malicious incoming traffic.

Organization

The rest of the research paper is organized as follows: The section literature review discusses the literature review of the previously proposed security approaches. After that, section proposed methodology discusses the proposed work with the working of ANFIS and the proposed security algorithm along with its method. Similarly, section experimental analysis and results discusses the implementation, results, and discussion. Moreover, lastly, section conclusion discusses the conclusion and future work.

LITERATURE REVIEW

According to previous research (1), healthcare-based networking servers require more security and a safe network architecture to provide confidentiality, integrity, privacy, and authentication to its patients, doctors, and management. The author proposed the ANFIS-based data classifier and security provider tool or model for making cloud servers more secure with a higher accuracy rate and smaller error rate. Detection of security attacks and malicious network traffic is considered as the most topical issue for network security (19–23). Therefore, previous research (24) presented an ANFIS-based security framework for classification of attacks and identifying their type.

In another study (25), the research mainly focuses on the interoperability of the cloud server platform and its reliability. The paper uses the hybrid approach of the squirrel search genetic algorithm with the combination of ANFIS to perform better functionality and remove uncertainties of servers by providing a higher accuracy rate. Also, in a previous work (26), an SDN anomaly detection system was proposed to detect malicious behavior and intrusion attacks using the cloud medium. The detection system identified the trusted edge for data or information sharing and communication. Based on multiple parameters, the implementation of the proposed systems showed better performance results. Here **Table 1** represents the ANFIS methodology used for security purposes like in clustering, classification, accuracy, performance evaluation, etc.

In many previous kinds of research, the communication between devices and multiple systems can also cause security threats and need efficient and enhanced network capacities during communication. Hence in a previous research (33), the author used 5G and the multiple-input multiple-output (MIMO) concept to enhance the network capacity and area coverage capacity. Similarly, in another paper (19), the author proposed the ANN technique to detect and identify intrusion attacks in android systems.

In a previous study (34), the author dealt with the security factor of the cloud to improve its performance using the AI technique. The proposed approach helped to overcome the flaw of data breaching and non-authorization access on cloud servers. They utilized different labels to restrict access to specific edge limits to ensure accuracy in decision making. It helped the systems to store a large amount of data securely on cloud servers.

TABLE 1 | Recent research on ANFIS and network security issues.

References	Type of risks	Approaches
Srilakshmi and Muthukuru (27)	Worm-hole and malicious nodes	Hybrid reactive search and bat (HRSB) mechanism used to detect malicious nodes and ANFIS for testing and training data
Pawar and Jagadeesan (28)	Black-hole attack	Used a self-adaptive multi-verse optimizer with ANFIS to detect intrusion attacks in WSN
Maheswari and Karthika (29)	Intrusion detection in network	ANFIS clustering methodology used for selecting cluster heads
Parfenov et al. (24)	Denial of service attack	ANFIS used to improve network traffic attack detection and performance evaluation
Nandi and Kannan (30)	Packet flooding attacks	ANFIS classifier used for feature extraction and classification in MANET
Hemalatha et al. (31)	Routing attack	ANFIS used for initial feature selection and trust evaluation
Barraclough et al. (32)	Phishing attacks	ANFIS-based classification approach for higher accuracy rate

TABLE 2 | Recent research on network security risks.

Types	Risks	Approaches
Data security	Unauthorized access, data leakage, data disclosure, privacy disclosure	Hybrid approach of the DSA algorithm with reverse flow (38). Hybrid encryption algorithm for securing big data storage (39).
Platform security	Data sharing, software, hardware, application	Combination of AES, Blowfish and Twofish security algorithms for secure data sharing (40). Enhanced role-based access algorithms to secure IoT-based cloud (41).
Application security	Configuring, system accessing, management	Hybrid framework of ECC and AES for advance encryption approach (42). Publisher subscriber algorithm with ontology logic for encryption to maintain confidentiality and authenticity (43).
Infrastructure security	Cloud framework, fault injection, false model, resource handling	Proposed RDFI strategies with chaos engineering algorithm (CEA) for deployed infrastructure (44). Used CSBAuditor to monitor and detect malicious attacks in the cloud environment (45).
Physical security	Sensitive data leakage, privacy breaching, hacking, intrusion	Lightweight cryptographic algorithms for physical security analysis (46). SHA-512, ECC, and RSA hybrid cloud security system used for securing data in a cloud structure (47).

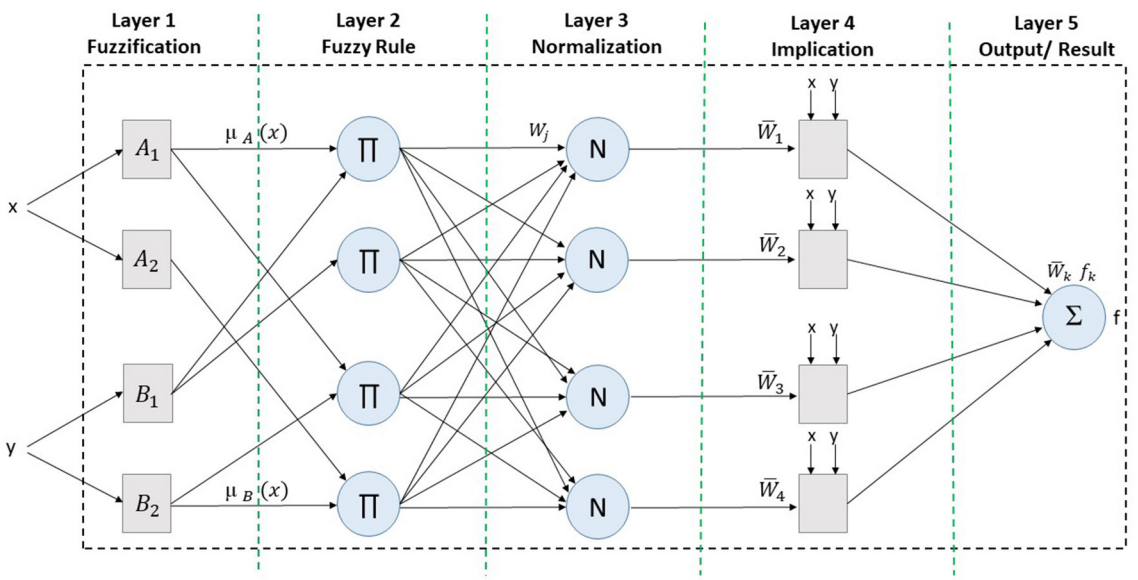


FIGURE 1 | Five layer-based working of the ANFIS architecture.

In an earlier paper (35), the study presented fuzzy-based detection schemes by various machine learning techniques and data mining algorithms to cope with multiple types of malicious attacks or intrusion. This paper first categorized their contribution in two parts, intrusion and detection, and then used fuzzy techniques for classification and identification. Then they

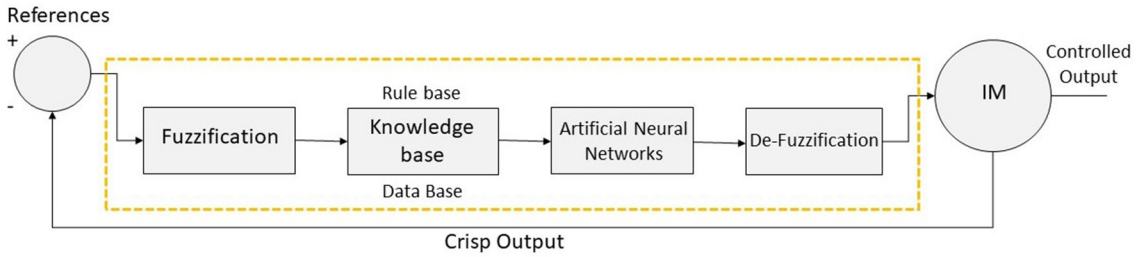


FIGURE 2 | The controller architecture of the ANFIS technique (ANFIS-C).

listed the shortcomings and merits of the ID detection based on fuzzy techniques and algorithms.

In another research paper (36), the author presented a detailed survey on anomaly detection schemes and fuzzy inference systems. This paper mainly focused on combining fuzzy inference systems and machine learning algorithms for the intrusion detection process. It summarized the research with the contributions and shed light on the shortcomings and benefits of FIS. Then, finally, it concluded with future findings and issues in the anomaly detection scheme.

CC is the computing solution that allows multiple end users to connect, share, communicate, and store data or information. Fault tolerance is considered the main concern in data reliability. Therefore, in a previous research (37), detailed analysis was presented to detect the nature of the error and proper implementation of the FIS technique to the response. The process of checkpoints was used to check the intensity of the error.

Table 2 represents the methodology used for cloud security based on previous research work.

In earlier research (48), a comprehensive review discussed CC and IoT in the healthcare sector including smart systems, smart applications, smart software, smart hospitals, and smart record systems. The study presented the IoT-based smart cloud paradigm in the healthcare sector. Similarly, In an earlier work (49), an ANFIS-based healthcare system was proposed to avoid network security risks. The proposed system and already existing ANFIS-based system were compared to evaluate the performance.

Based on previous work, ANFIS is considered an advanced technique for data evaluation, classification, clustering, increased accuracy rate, and detecting various network or systems attacks or any malicious activities on cloud infrastructure to provide a trustful environment to its end users.

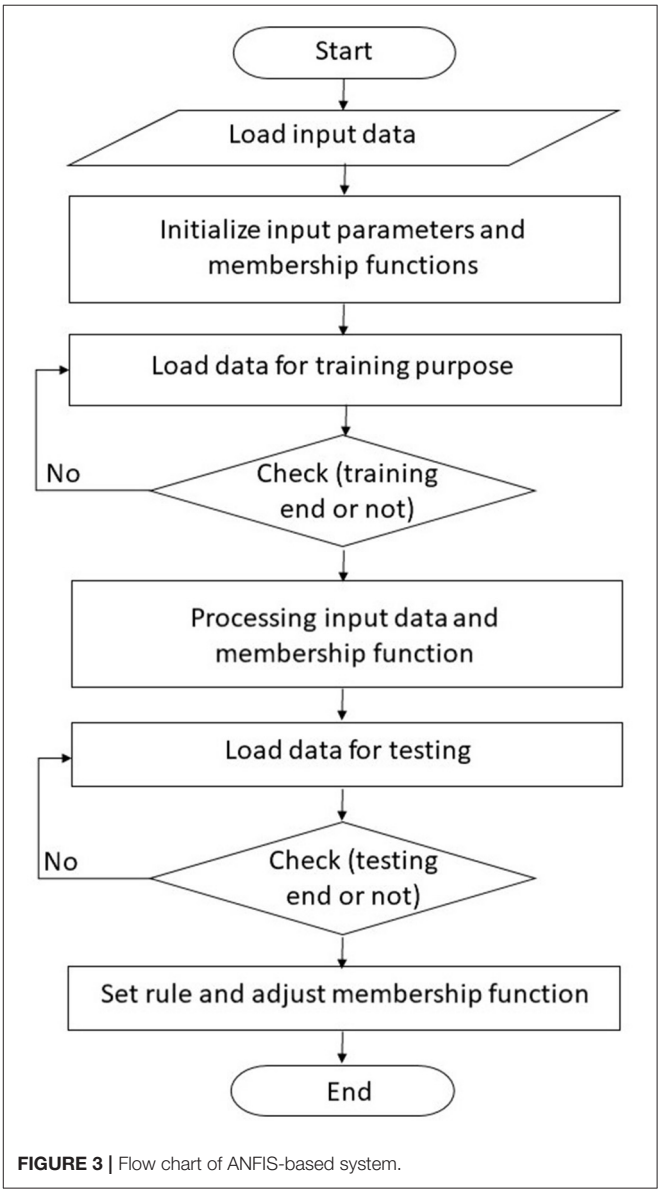


FIGURE 3 | Flow chart of ANFIS-based system.

PROPOSED METHODOLOGY

We propose a tuned version of the ANFIS model to detect intrusion attacks in cloud database servers. We mainly focus on identifying the unauthorized access of end users to systems and diagnosing the type of network attack. The proposed security algorithm will help systems to detect malicious activities or irrelevant information from network traffic. The coming subsection will help understand the ANFIS model's working and the proposed technique's flow chart.

Architecture and Working of ANFIS-Based Systems

To study fuzzy rules-based non-linear systems, the fuzzy inference system (FIS) is considered an efficient methodology. Similarly, artificial neural networks (ANNs) work on neurons and the artificial Intelligence technique. So, an artificial neural fuzzy inference system (ANFIS) is the combination of FIS and ANN, which works on the principle of neural networks and follows the efficiency of FIS. ANFIS uses the hybrid Sugeno-type methodology to specify the input parameters. This method trains the membership function based on input parameters to get the trained dataset. ANFIS is also known for its the parameters which can validate the model. ANFIS is based on a multiple layer architecture, where every layer keeps forwarding the input parameters for continuous working. More detail about the architecture and working of the ANFIS architecture is presented in **Figure 1**.

The architecture of the ANFIS methodology almost resembles that of the Sugeno-type model, as shown in **Figure 1**. The model-based rule sets are as follow:

- If x is A_1 and y is B_1 then f_1 = p_1 x + q_1 y + r_1
- If x is A_2 and y is B_2 then f_2 = p_2 x + q_2 y + r_2

The working of each layer is explained with mathematical equations as listed below:
Layer 1 shows the input parameters and fuzzification, where Ni (N is node) is adaptive to N func (func is function), as represented in Equations 1 and 2.

$$L_1, i = \mu_A i(x), for i = 1, 2 \tag{1}$$

$$L_1, i = \mu_B i - 2(x), for i = 3, 4 \tag{2}$$

Layer 2 shows the fuzzy rule generation phase, where every N computes and sets rules by the multiplication process, as represented in Equation 3.

$$L_2, i = w_i = \mu_A i(x).\mu_B i(y), i = 1, 2 \tag{3}$$

Layer 3 shows the normalization phase, where every neuron is normalized based on the effect of the fuzzy rule sets, as represented in Equation 4.

$$L_3, i = \bar{w} = \frac{w_i}{w_1 + w_2}, i = 1, 2 \tag{4}$$

Layer 4 shows the implication phase, where every input value is set as an input parameter, as represented in Equation 5.

$$L_4, i = \bar{w_i} f_i = \bar{w_i}(p_i x + q_i y + r_i), i = 1, 2 \tag{5}$$

Layer 5 shows the final result or output of the complete process, where the sum of the input values are computed, as represented in Equation 6.

$$L_5, i = \Sigma \bar{w} f_i = \frac{\Sigma w_i f_i}{\Sigma w_i} \tag{6}$$

TABLE 3 | Input/Output parameters and membership functions.

Types of attacks	Ranges and rules
[Input 1] [Normal]	Range: [0 1], MFs: 3 (low, medium, high)
[Input 2] [Probe]	Range: [0 1], MFs: 3 (low, medium, high)
[Input 3] [DoS]	Range: [0 1], MFs: 3 (low, medium, high)
[Input 4] [U2R]	Range: [0 1], MFs: 3 (low, medium, high)
[Input 5] [R2L]	Range: [0 1], MFs: 3 (low, medium, high)
[Output] [Attack_type]	Range: [0 1], MFs: 3 (low, medium, high)

Figure 2 shows the ANFIS-based controller, which helps the systems detect attacks (ANFIS-C). The controller helps to detect the type of attack based on input parameters and membership function. Through the learning process, ANFIS-C adjusts the initial and processed parameters. The least-square method is used here to obtain hybrid propagation. The ANFIS-C alters the initial and processed parameters based on the error rate (here, error rate is calculated with a difference of two output values before processing and after ANFIS-C).

Figure 3 shows the working of the ANFIS-based intelligence system. The initial step involves loading input parameters, and the ANFIS-based classifier evaluates the functions and detects the error rate, intelligence method, and other learning attributes. After that, the training phase defines the initial and processed membership function along with its parameters. Then, after passing from the testing phase, it sets rules and defines membership functions. At last, the ANFIS-based system stores processed parameters and processed attributes.

EXPERIMENTAL ANALYSIS AND RESULTS

The ANFIS model is used to detect the type of attack based on rule viewer, membership function, and surface viewer. For the practical implementation of the ANFIS model, we use the MATLAB framework for experiments and results. We use dataset pf KDDcup 99 for intrusion detection. This dataset consists of 41 features of the network and five network types of intrusions. The network is observed for around 7 weeks to collate more accurate and precise data. We use an ANFIS-based classifier for detection purposes. The basic structure and features of the ANFIS model in PHR network attacks can be examined. In **Table 3** input parameters and membership functions are discussed.

Similarly, **Figure 4** shows the five input parameters (based on attacks type): normal, probe, DoS, U2R, and R2L.

On the basis of input parameters, membership function is defined using the Sugeno-type ANFIS function in **Figure 5**. More than 140 rules are generated to generate the desired output value in the form of attack type. After selecting various combinations of input parameters, various rules are generated to get the results.

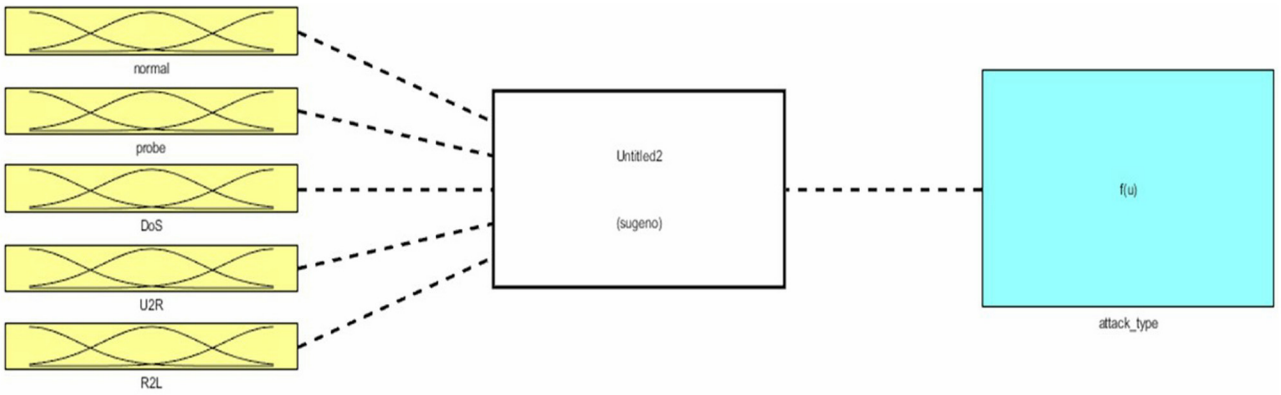


FIGURE 4 | Input parameters.

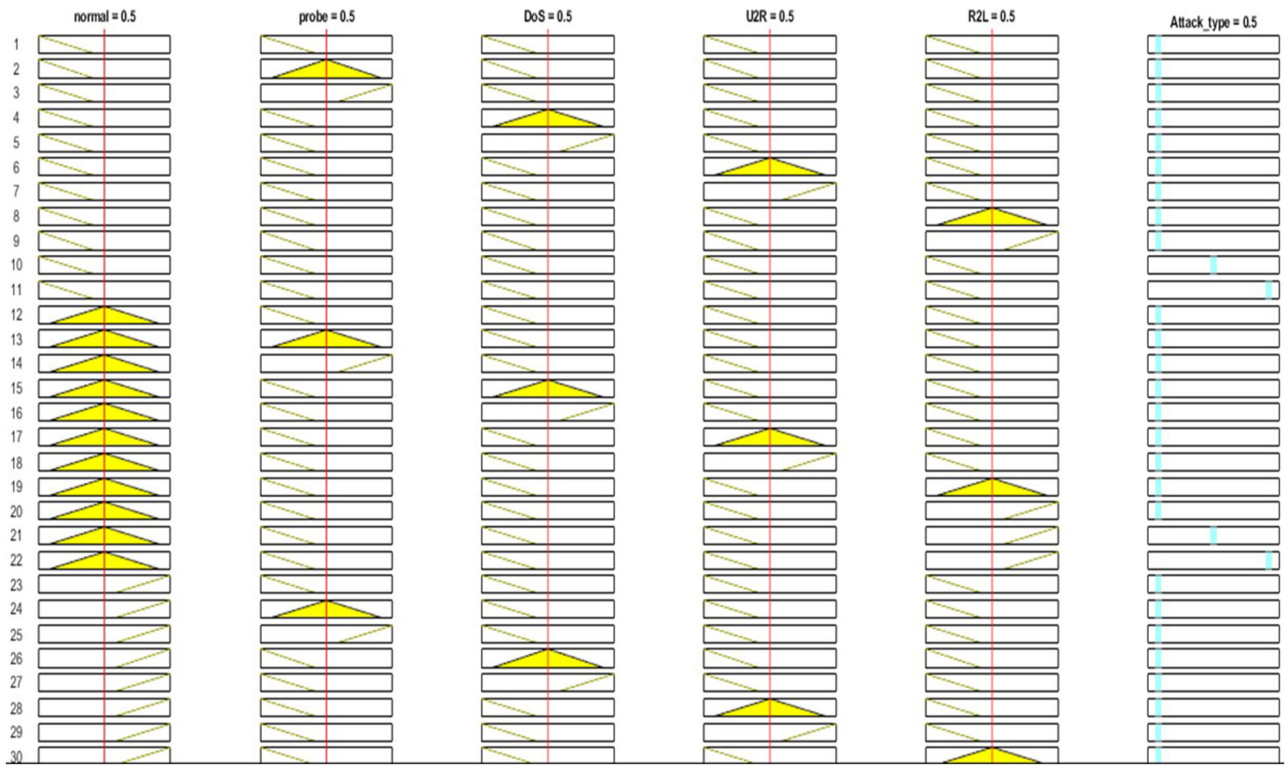


FIGURE 5 | Set of rules based on membership functions.

Figure 6 shows the 3D surface views of all attributes. In section (**Figure 6A**), for attack type the x-axis displays "normal" and the y-axis shows "probe"; then in section (**Figure 6B**), the x-axis presents "probe" and the y-axis shows "normal"; in section (**Figure 6C**), the x-axis shows "U2R" and the y-axis shows "normal"; then in section (**Figure 6D**), the x-axis displays "DoS" and the y-axis presents "normal"; and lastly in section (**Figure 6E**), the x-axis displays "R2L" and the y-axis shows "normal." The surface view shows the comparison of multiple attack type attributes. As the KDD99cup dataset consists of many types of network attacks. In our research, we choose five types of attacks based on their repetition. Moreover, the surface viewer helps predict the accuracy rate in classification and detection based on a rule (which is created by using the membership function).

Similarly, after defining rule sets, the structure of the proposed ANFIS-based model is established, as can be seen in **Figure 7**. In the final structure, after multiple input values, around 140 sets

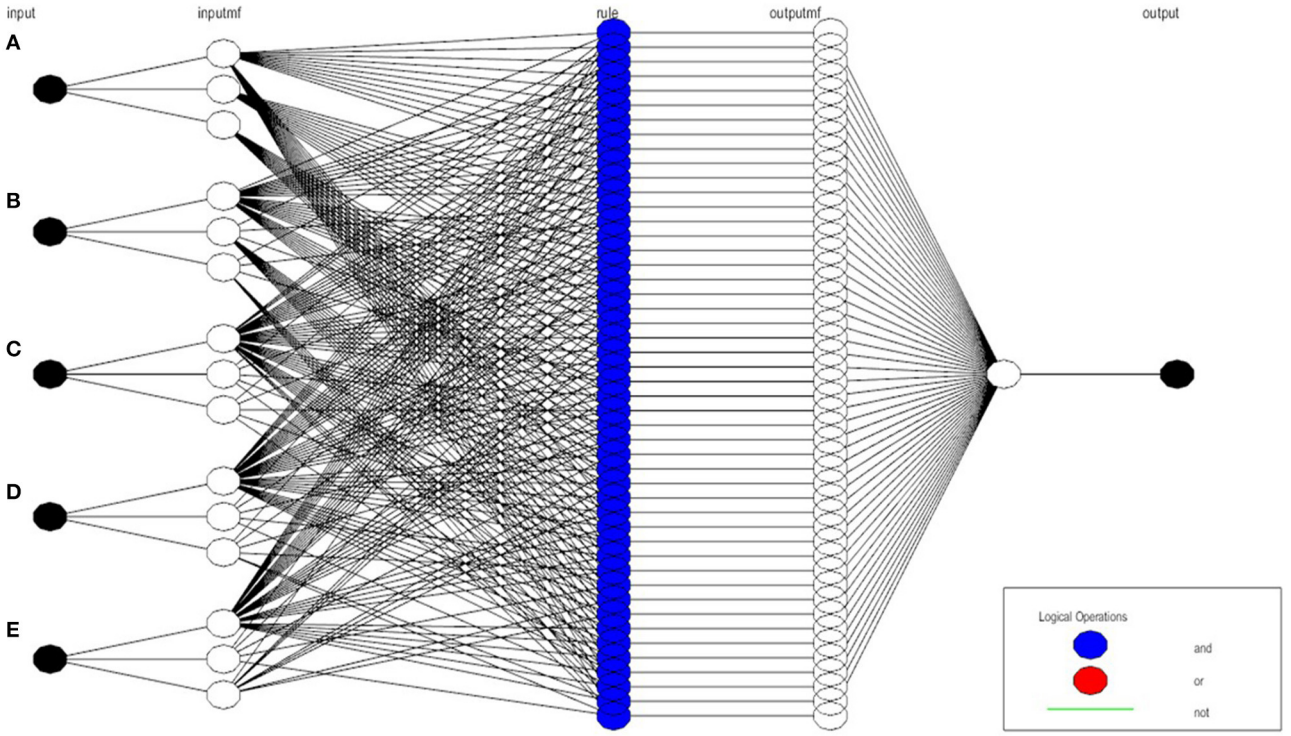


FIGURE 6 | (A–E) Set of rules based on membership functions.

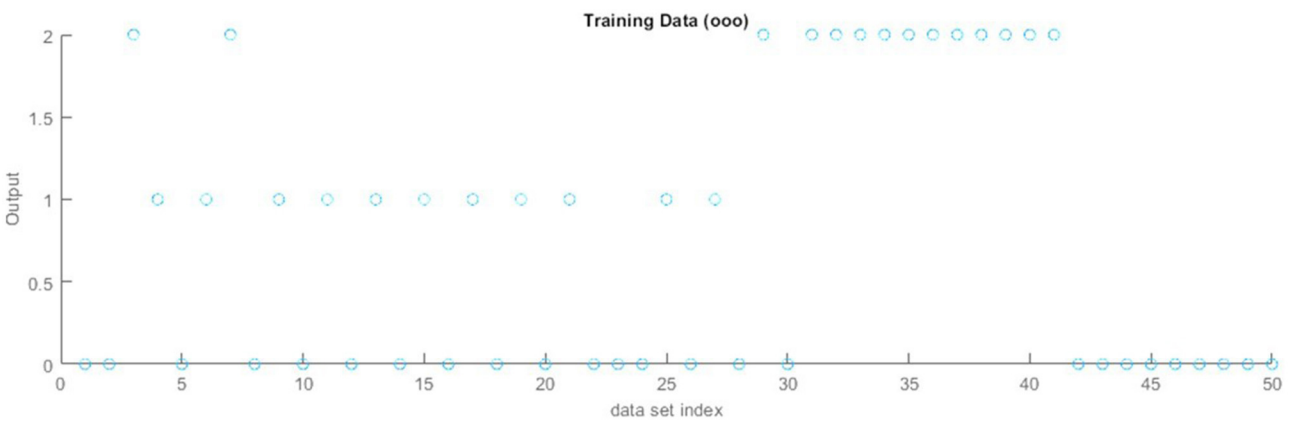


FIGURE 7 | Proposed ANFIS model.

of rules are generated which leads toward the final output value, which indicates the type of intrusion attack.

Figure 8 represents the training results of the KDDcup 99 datasets collected from the Kaggle website for network intrusion detection based on ANFIS. The training model contains five input values and generates a single output. The training model shows the increasing flow of accuracy rate.

Finally, **Table 4** shows the detailed statistical result of the KDDcup 99 datasets. The selected classes for testing and training are normal, probe, DoS, R2L, and U2R.

The final result generated by the ANFIS model is either linear or non-linear. Therefore, the Sugeno-type ANFIS model helps to obtain a single output. The results show that the PHR in the healthcare system require an ANFIS-type framework to detect and prevent network attacks and maintain secrecy and a trustful

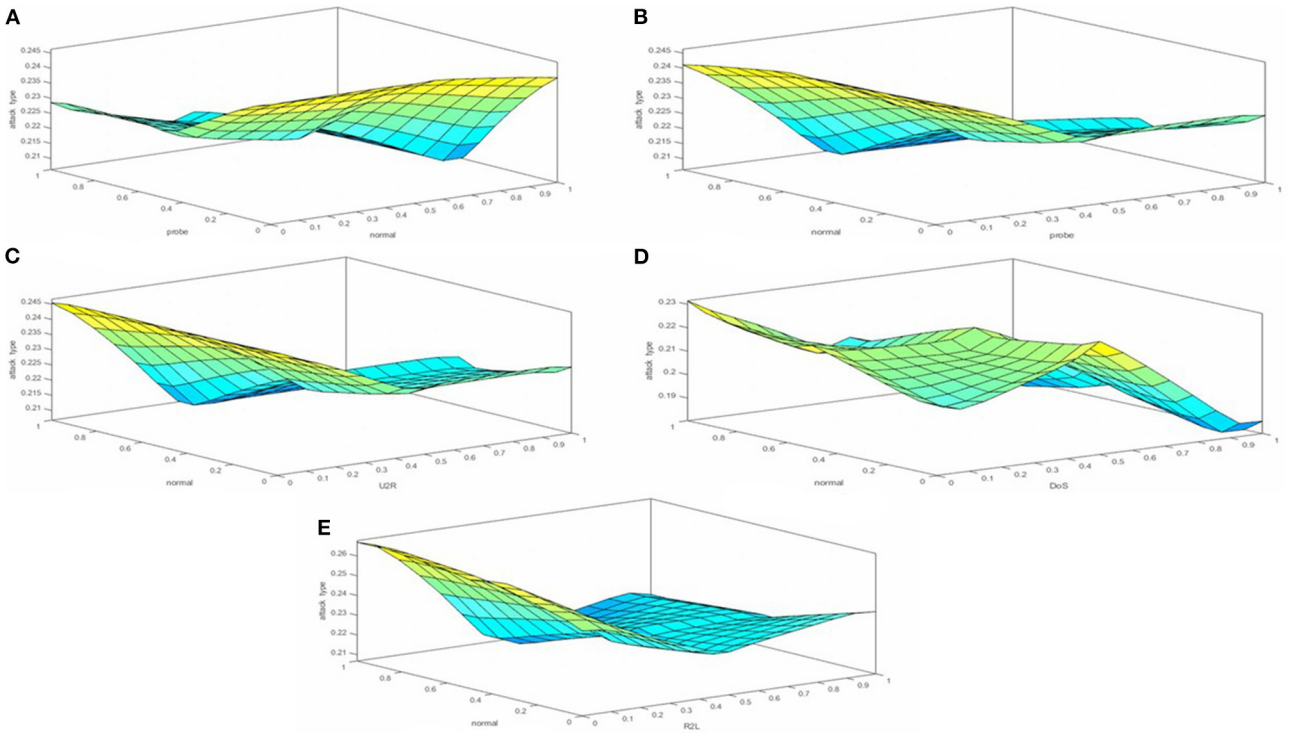


FIGURE 8 | Training results of the KDDcup 99 dataset. **(A)** On x-axis "normal". **(B)** On x-axis "probe". **(C)** On x-axis "U2R". **(D)** On x-axis "DoS". **(E)** On x-axis "R2L".

TABLE 4 | Testing and training results from the KDDcup 99 datasets.

Types of attacks	Details	Training result	Testing result
Normal	Stable connectivity	97,278	60,593
Probe	Configuration and analysis details of the system and network	4,107	4,166
DoS	Affecting network resources	391,458	229,853
U2R	Accessibility to servers and connected nodes	52	258
R2L	Illegal accessibility to remote devices	1,126	16,189

TABLE 5 | Comparison of FIS and ANFIS models.

Performance	FIS	ANFIS (Proposed)
MSE	0.0183	0.0123
NMSE	0.3185	0.2650
MAE	0.1170	0.0747
Error (min. obs)	0.0110	0.0021
Error (max. obs)	0.1279	0.1706
R-value	0.6133	0.7336

environment for patients, doctors, laboratories, and all other interlinked sectors.

Comparison of Experimented FIS and ANFIS Model Decision-Making

Table 5 shows the performance comparison of the FIS and ANFIS models for better decision-making in network intrusion detection to make e-healthcare (PHR) more secure. The FIS model has been selected to conduct another experiment to show the comparison of the results of FIS and ANFIS based on performance. The mentioned table helps to evaluate the performance of both models. Other than selecting already proposed solutions, we train the FIS model on the same datasets to obtain results. The computed MSE value for FIS is 0.0183 and 0.0123 for ANFIS, the NMSE value is 0.3185 for FIS and 0.2650 for ANFIS, the MAE value is 0.1170 for FIS and 0.0747 for ANFIS, the error value (for minimum observation) is 0.0110 for FIS and 0.0021 for ANFIS, the error value (for maximum observation) is 0.1279 for FIS and 0.1706 for ANFIS, and then finally the R-value is 0.6133 for FIS and 0.7336 for ANFIS. Based on the performance results and output values, ANFIS is considered a more efficient model for efficient decision-making in e-healthcare systems.

CONCLUSION

Security of network systems in all sectors has become one of the most crucial challenges for researchers. In smart e-healthcare systems, cloud servers provide a better intermediary platform for data storage, communication, and many other aspects. The end users (doctors, clinical experts, patients, laboratories) get the

opportunity to access PHR and access the networking database servers. It can be any authorized person. It is essential to detect and classify malicious activities and network traffic. This research paper proposed ANFIS for effective intrusion detection in healthcare networks to maintain a patient record. The main agenda is to detect the unauthorized access of users by using an ANFIS-based intrusion detection system. The proposed security algorithm helps to collect malicious activities or information from network traffic. Based on the if-then rule statement and ANFIS-based data classification that helps to detect the intrusion attack, it can determine whether the networking database server was hit by an intruder or not. If the system identifies the type of attack, this can help researchers put precautionary measures into place to overcome the loss or block the malicious incoming traffic.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

FA and DL: conceptualization. DL: data curation. FA: formal analysis, investigation, and methodology. NK: funding acquisition. NK and PZ: project administration. SA and PZ: resources. FA and MR: software. FA and NK: supervision. DL and SA: validation. DL and NK: visualization. PZ and SA: writing–review and editing. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Mohiyuddin A, Javed AR, Chakraborty C, Rizwan M, Shabbir M, Nebhen J. Secure cloud storage for medical IoT data using adaptive neuro-fuzzy inference system. *Int J Fuzzy Sys*. (2021). doi: 10.1007/s40815-021-01104-y
2. Javed A, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. Pp-spa: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Proc Lett*. (2021) 1–18. doi: 10.1007/s11063-020-10414-5
3. Javed AR, Abid R, Aslam B, Khalid HA, Khan MZ, Alhazmi Rehman SU, et al. Personalisedcomfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents. *Enterprise Inf Syst*. (2020) 1-23. doi: 10.1080/17517575.2020.1852316
4. Shabbir M, Shabbir A, Iwendi C, Javed AR, Rizwan M, Herencsar N, et al. Enhancing security of health information using modular encryption standard in mobile cloud computing. *IEEE Access*. (2021) 9:8820-34. doi: 10.1109/ACCESS.2021.3049564
5. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an IPFS-based optimization algorithm. *J Circ Syst Comput*. (2021) 2250010. doi: 10.1142/S0218126622500104
6. Garg D, Sidhu J, Rani S. A note on cloud computing security. Int J Ad Hoc Ubiquit Comput. (2020) 3:133-54. doi: 10.1504/IJAHUC.2020.106644
7. Ikram AA, Rehman A, Rizwan M, Abid R, Crichigno J, Srivastava G. Mobile cloud computing framework for securing data. In: 2021 44th International Conference on Telecommunications and Signal Processing (TSP) Brno: IEEE (2021). p. 309-15.
8. Wang Z, Wang N, Su X, Ge S. An empirical study on business analytics affordances enhancing the management of cloud computing data security. *Int J Inf Manag*. (2020) 50:387–394. doi: 10.1016/j.ijinfomgt.2019.09.002
9. Naeem A, Rehman A, Rizwan M, Abbas S, Lin JCW, Gadekallu TR. DARE-SEP: a hybrid approach of distance aware residual energy-efficient SEP for WSN. *IEEE Trans Green Commun Netw*. (2021) 5:611–621. doi: 10.1109/TGCN.2021.3067885
10. Alabdulatif A, Khalil I, Yi X. Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption. *J Parallel Distrib Comput*. (2020) 137:192–204. doi: 10.1016/j.jpdc.2019.10.008
11. Yousuf H, Lahzi M, Salloum SA, Shaalan K. Systematic review on fully homomorphic encryption scheme and its application. In: *Recent Advances in Intelligent Systems and Smart Applications*. Cham: Springer. (2021) p. 537–51.
12. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inform*. (2021). doi: 10.1109/JBHI.2021.3112693. [Epub ahead of print].
13. Butt UA, Mehmood M, Shah SBH, Amin R, Shaukat MW, Raza SM, et al. A review of machine learning algorithms for cloud computing security. *Electronics*. (2020) 9:1379. doi: 10.3390/electronics9091379
14. Sabu S, Ramalingam H, Vishaka M, Swapna H, Hegde S. Implementation of a Secure and privacy-aware E-Health record and IoT data Sharing using Blockchain. *EasyChair*. (2021) doi: 10.1016/j.gltp.2021.08.033
15. Muhammad A, Asad M, Javed A. Robust early stage botnet detection using machine learning. In: *2020 International Conference on Cyber Warfare and Security (ICCWS)*. Islamabad: IEEE (2020). p. 1–6.
16. Basit A, Zafar M, Javed AR, Jalil Z. A novel ensemble machine learning method to detect phishing attack. In: *2020 IEEE 23rd International Multitopic Conference (INMIC)*. Bahawalpur: IEEE (2020). p. 1-5.
17. Basit A, Zafar M, Liu X, Javed A, Jalil Z, Kifayat K. A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommun Syst*. (2021) 76:139–54. doi: 10.1007/s11235-020-00733-2
18. Abid R, Iwendi C, Javed AR, Rizwan M, Jalil Z, Anajemba JH, et al. An optimised homomorphic CRT-RSA algorithm for secure and efficient communication. *Pers Ubiquitous Comput*. (2021) 1–14. doi: 10.1007/s00779-021-01607-3
19. Imtiaz SI, ur Rehman S, Abdul R, Jalil Z, Liu X, Alnumay WS. DeepAMD: detection and identification of android malware using high-efficient deep artificial neural network. *Future Gen Comput Syst*. (2021) 115:844–56. doi: 10.1016/j.future.2020.10.008
20. Ahmed W, Rasool A, Abdul R, Kumar N, Gadekallu TR, Jalil Z, et al. Security in next generation mobile payment systems: a comprehensive survey. *IEEE Access*. (2021) 9:115932–50. doi: 10.1109/ACCESS.2021.3105450
21. Mittal M, Iwendi C, Khan S, Rehman A. Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using Levenberg-Marquardt neural network and gated recurrent unit for intrusion detection system. *Trans Emerg Telecommun Technol*. (2021) 32:e3997. doi: 10.1002/ett.3997
22. Iwendi C, Rehman SU, Javed AR, Khan S, Srivastava G. Sustainable security for the internet of things using artificial intelligence architectures. *ACM Trans Intern Technol*. (2021) 21:1–22. doi: 10.1145/3448614
23. Afzal S, Asim M, Abdul R, Beg MO, Baker T. URLdeepDetect: a deep learning approach for detecting malicious urls using semantic vector models. *J Netw Syst Manag*. (2021) 29:1–27. doi: 10.1007/s10922-021-09587-8
24. Parfenov D, Zabrodina L, Bolodurina I, Parfenov A. Development of Algorithmic Solutions for Solving the Problem of identifying Network Attacks Based on Adaptive Neuro-Fuzzy Networks ANFIS. In: *2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. Yekaterinburg: IEEE (2020).
25. Ramalingam C, Mohan P. An efficient applications cloud interoperability framework using I-Anfis. *Symmetry*. (2021) 13:268. doi: 10.3390/sym13020268
26. Qureshi KN, Jeon G, Piccialli F. Anomaly detection and trust authority in artificial intelligence and cloud computing. *Comput Netw*. (2021) 184:107647. doi: 10.1016/j.comnet.2020.107647

27. Srilakshmi R, Muthukuru J. Intrusion detection in mobile ad-hoc network using hybrid reactive search and bat algorithm. *Int J Intell Unmanned Syst*. (2021) doi: 10.1108/IJIUS-09-2020-0049
28. Pawar MV, Jagadeesan A. Detection of blackhole and wormhole attacks in WSN enabled by optimal feature selection using self-adaptive multi-verse optimiser with deep learning. *Int J Commun Netw Distrib Syst*. (2021) 26:409–45. doi: 10.1504/IJCNDS.2021.115573
29. Maheswari M, Karthika R. A novel QoS based secure unequal clustering protocol with intrusion detection system in wireless sensor networks. *Wireless Pers Commun*. (2021) 118:1535–57. doi: 10.1007/s11277-021-08101-2
30. Nandi M, Kannan A. An optimized and hybrid energy aware routing model for effective detection of flooding attacks in a manet environment. (2021). doi: 10.21203/rs.3.rs-586844/v1
31. Hemalatha R, Umamaheswari R, Jothi S. ANFIS based optimal routing using group teaching and adaptive equilibrium optimization based trust aware routing protocol in MANET. (2021). doi: 10.21203/rs.3.rs-355720/v1
32. Barraclough PA, Fehringer G, Woodward J. Intelligent cyber-phishing detection for online. *Comput Secur*. (2021) 104:102123. doi: 10.1016/j.cose.2020.102123
33. Javed AR, Abid R, Aslam B, Khalid HA, Khan MZ, Alhazmi OH, et al. Green5g: enhancing capacity and coverage in device-to-device communication. *Comput Mater Continua*. (2021) 67:1933–50. doi: 10.32604/cmc.2021.015272
34. Bhattacharya D, Biswas A, Rajkumar S, Selvanambi R. Dynamic cloud access security broker using artificial intelligence. In: *Machine Learning for Predictive Analysis*. Springer (2021). p. 335–42.
35. Masdari M, Khezri H. A survey and taxonomy of the fuzzy signature-based intrusion detection systems. *Appl Soft Comput*. (2020) 106301. doi: 10.1016/j.asoc.2020.106301
36. Masdari M, Khezri H. Towards fuzzy anomaly detection-based security: a comprehensive review. *Fuzzy Optimizat Decis Mak*. (2021) 20:1–49. doi: 10.1007/s10700-020-09332-x
37. Rezaeipanah A, Mojarad M, Fakhari A. Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic. *Int J Comput Appl*. (2020) 1–9. doi: 10.1080/1206212X.2019.1709288
38. Ferdous J, Khan MFN, Rezaul KM, Tamal MA, Aziz MA, Miah P. A hybrid framework for security in cloud computing based on different algorithms. *Int J Netw Secur*. (2020) 22:638–44.
39. Viswanath G, Krishna PV. Hybrid encryption framework for securing big data storage in multi-cloud environment. *Evol Intell*. (2020) p. 1-8. doi: 10.1007/s12065-020-00404-w
40. Joel MR, Ebenezer V, Navaneethakrishnan M, Karthik N. Encrypting and decrypting different files over different algorithm on Cloud Platform. *Int J Emerg Trends Eng Res*. (2020) 8. doi: 10.30534/ijeter/2020/71842020
41. Ramesh S, Jayasankar T, Bhavadharini R, Nagarajan N, Mani G. Securing medical data using extended role based access control model and twofish algorithms on cloud platform. *Eur J MolClin Med*. (2021) 8:1075–89.
42. Chavan S, Tamane S. Enhancement in Cloud security for web application attacks. In: *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*. Aurangabad: IEEE (2020). p. 91–95.
43. Orobosade A, Aderonke T, Boniface A, Gabriel AJ. Cloud application security using hybrid encryption. *Communications*. (2020) 7:25–31. doi: 10.5120/cae2020652866
44. Torkura KA, Sukmana MI, Cheng F, Meinel C. Cloudstrike: chaos engineering for security and resiliency in cloud infrastructure. *IEEE Access*. (2020) 8:123044–60. doi: 10.1109/ACCESS.2020.3007338
45. Torkura KA, Sukmana MI, Cheng F, Meinel C. Continuous auditing and threat detection in multi-cloud infrastructure. *Comput Secur*. (2021) 102:102124. doi: 10.1016/j.cose.2020.102124
46. Pallavi K, Kumar VR, Srikrishna S. Comparative study of various lightweight cryptographic algorithms for data security between IoT and cloud. In: *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. Coimbatore: IEEE (2020). p. 589–93.
47. Soman VK, Natarajan V. Analysis of hybrid data security algorithms for cloud. In: *Second International Conference on Networks and Advances in Computational Technologies*. Springer (2021). p. 231–42.
48. Darwish A, Hassanien AE, Elhoseny M, Sangaiah AK, Muhammad K. The impact of the hybrid platform of internet of things and cloud computing on healthcare systems: opportunities, challenges, and open problems. *J Ambient Intell Humaniz Comput*. (2019) 10:4151–66. doi: 10.1007/s12652-017-0659-1
49. Kaur J, Khan AI, Abushark YB, Alam MM, Khan SA, Agrawal A, et al. Security risk assessment of healthcare web application through adaptive neuro-fuzzy inference system: a design perspective. *Risk Manag Healthc Policy*. (2020) 13:355. doi: 10.2147/RMHP.S233706

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors MR.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Akram, Liu, Zhao, Kryvinska, Abbas and Rizwan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fpubh.2021.781827

Neural Network Based Mental Depression Identification and Sentiments Classification Technique From Speech Signals: A COVID-19 Focused Pandemic Study

Syed Thouheed Ahmed[1], *Dollar Konjengbam Singh*[2], *Syed Muzamil Basha*[3], *Emad Abouel Nasr*[4]*, *Ali K. Kamrani*[5] and *Mohamed K. Aboudaif*[4]*

[1] School of Computing and Information Technology, REVA University, Bengaluru, India, [2] ICT for Internet and Multimedia, University of Padua, Padua, Italy, [3] School of Computer Science and Engineering, REVA University, Bengaluru, India, [4] Industrial Engineering Department, College of Engineering, King Saud University, Riyadh, Saudi Arabia, [5] Industrial Engineering Department, College of Engineering, University of Houston, Houston, TX, United States

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Parvaiz Ahmad Naik,
Xi'an Jiaotong University, China
Ebuka Ibeke,
Robert Gordon University,
United Kingdom

***Correspondence:**
Emad Abouel Nasr
eabdelghany@ksu.edu.sa
Mohamed K. Aboudaif
maboudaif@ksu.edu.sa

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 23 September 2021
Accepted: 08 November 2021
Published: 06 December 2021

Citation:
Ahmed ST, Singh DK, Basha SM, Abouel Nasr E, Kamrani AK and Aboudaif MK (2021) Neural Network Based Mental Depression Identification and Sentiments Classification Technique From Speech Signals: A COVID-19 Focused Pandemic Study. Front. Public Health 9:781827. doi: 10.3389/fpubh.2021.781827

COVID-19 (SARS-CoV-2) was declared as a global pandemic by the World Health Organization (WHO) in February 2020. This led to previously unforeseen measures that aimed to curb its spread, such as the lockdown of cities, districts, and international travel. Various researchers and institutions have focused on multidimensional opportunities and solutions in encountering the COVID-19 pandemic. This study focuses on mental health and sentiment validations caused by the global lockdowns across the countries, resulting in a mental disability among individuals. This paper discusses a technique for identifying the mental state of an individual by sentiment analysis of feelings such as anxiety, depression, and loneliness caused by isolation and pauses to the normal chains of operations in daily life. The research uses a Neural Network (NN) to resolve and extract patterns and validate threshold trained datasets for decision making. This technique was used to validate 2,173 global speech samples, and the resulting accuracy of mental state and sentiments are identified with 93.5% accuracy in classifying the behavioral patterns of patients suffering from COVID-19 and pandemic-influenced depression.

Keywords: sentiment extraction, speech signal processing, COVID-19, mental depression, neural network

INTRODUCTION

The world is at present facing an uncertain time due to the global pandemic caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), also known as COVID-19. The pandemic has forced nations to exercise lockdown as a preventive measure to slow the spread of the virus. The pandemic has resulted in economic failure and disruptions in the supply chain all over the world. There was a race for a vaccine among modern drug and research organizations. The pandemic has caused major adverse effects such as mental depression, isolation, anxiety, and loneliness besides respiratory disorders and aligned symptoms. Depression and other mental health issues have been caused by lockdown and restrictions in travel and work, with a new normal social life now conducted via technological platforms.

The pandemic has bought a sense of maturity and adverse implications concerning psychosocial behaviors and mental health implications such as depression, anxiety, and loneliness. In this research, a systematic evaluation was conducted on the behavior of users based on speech signals, which were recorded using a machine learning technique to extract keywords and classify data using sentimental analysis techniques (1). The research in this article also focuses on identifying the user's mental state via the speech signals recorded over technological platforms used for virtual meetings and other gatherings (2). The research aims to provide schematic evaluation and validation approaches and classify patients based on medical conditions.

LITERATURE SURVEY

The global pandemic situation under COVID-19 has left traces of various adverse effects on people, which are the result of isolation, lockdown, mental health destabilization, and much more. The authors Singh et al. (3) and Naik et al. (4) have discussed the impact caused by COVID-19 and lockdown on the younger generation, focussing on children's behavior and reactions to the new normal. The study shows the overall implications of isolation on children and adolescents. Pfefferbaum and North (5) have discussed the impact and relationship of mental stress caused due to the global pandemic situation, with detailed insights into public health emergencies and the influence of the pandemic on looming health conditions. Furthermore, a discussion on the challenges faced by health care workers (HCW) and their state of mental stress is documented by Spoorthy et al. (6). The HCWs are frontline attributes and hence require assistance in evaluating and validating mental health via the main mode of communication now used, i.e., speech signals through digital media platforms and applications and a similar discussion is highlighted in other studies (7, 8).

Some studies are focused on the terms of technological solutions for the mental distress caused by the pandemic. These solutions have outlined the use of a telemedicine approach for reaching the maximum and remote population of a developing country like India. A study by Ahmed et al. (9) discussed

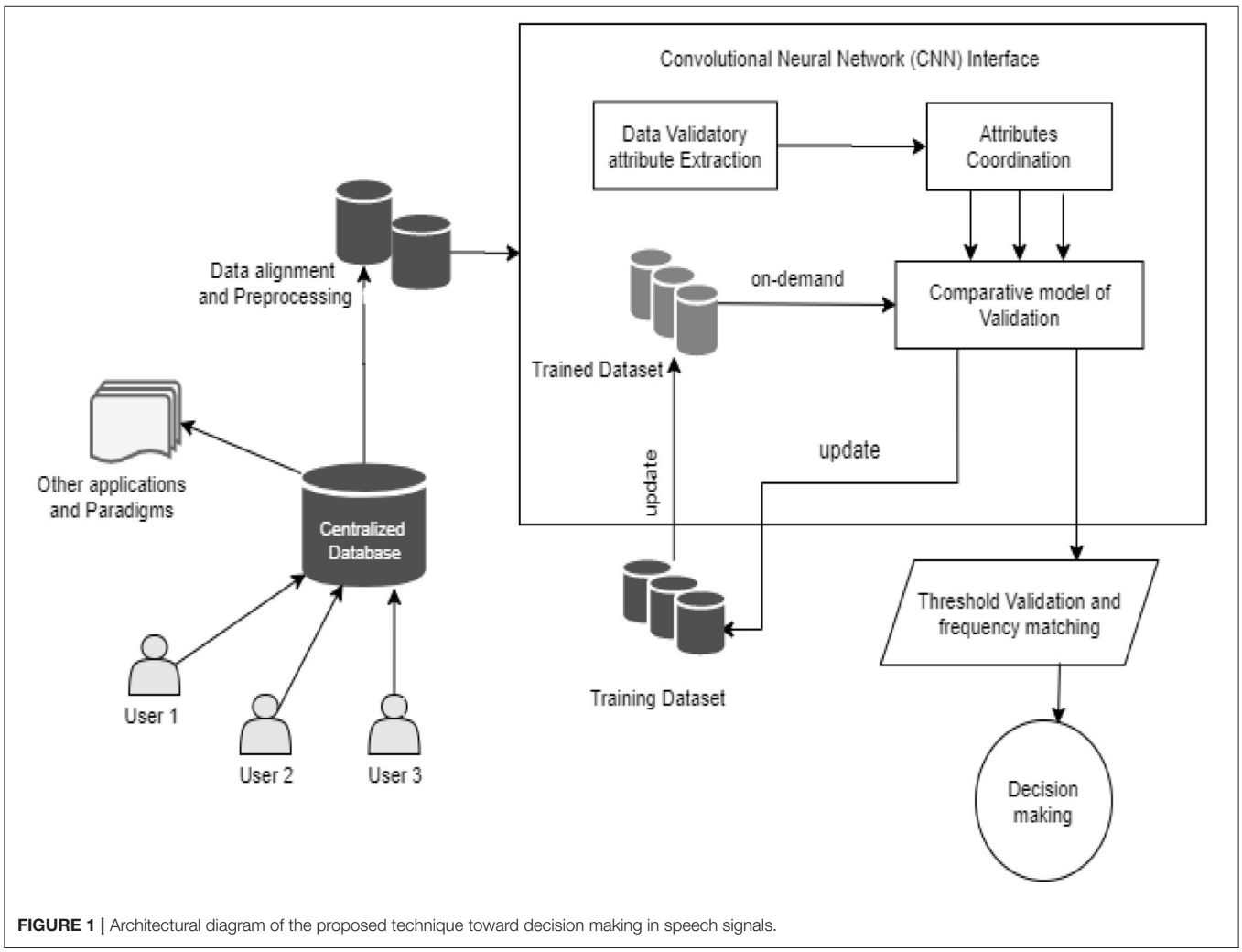


FIGURE 1 | Architectural diagram of the proposed technique toward decision making in speech signals.

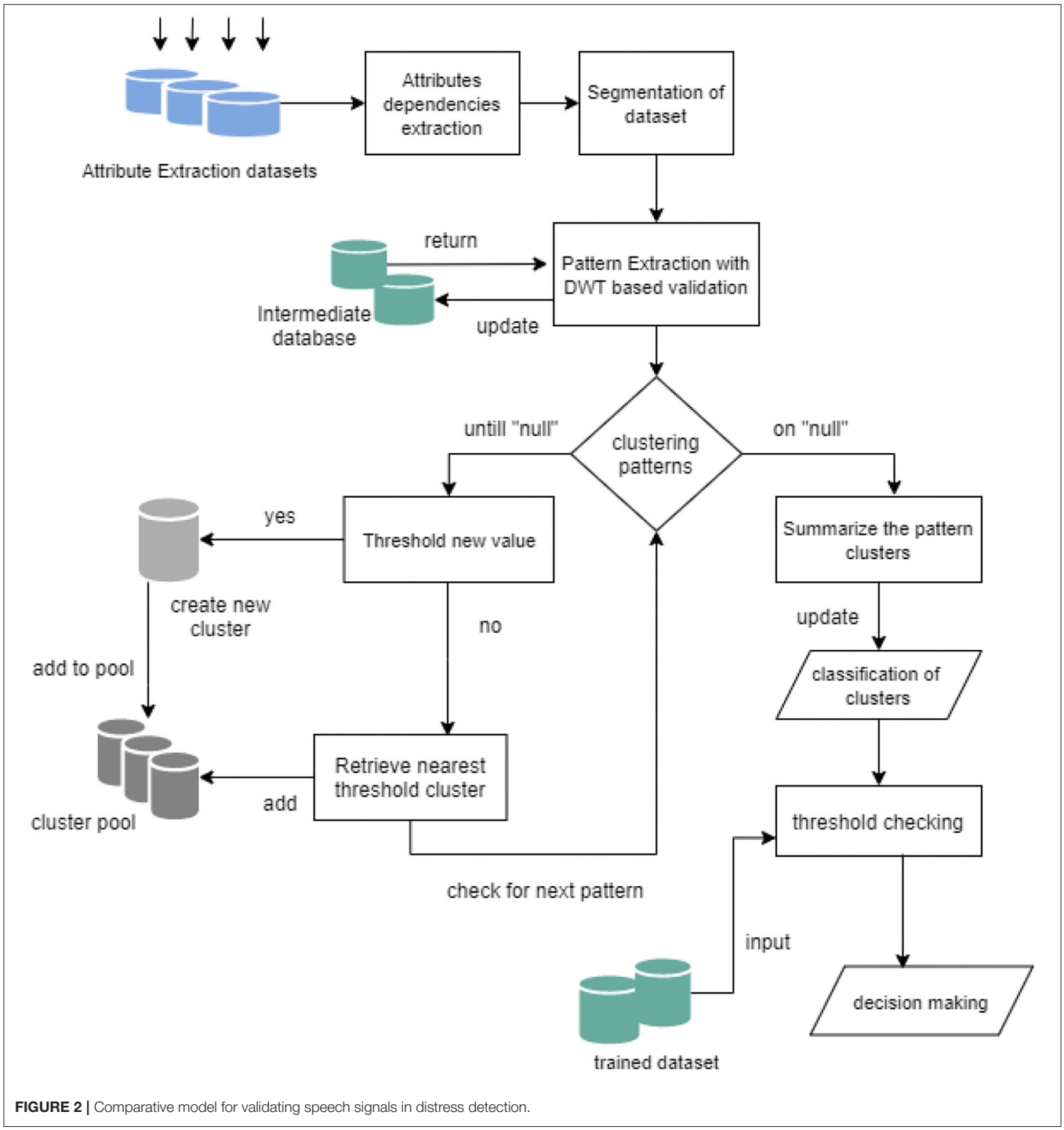


FIGURE 2 | Comparative model for validating speech signals in distress detection.

Multidimensional Optimal Medical Dataset processing under a telemedicine channel. These MooM datasets include a signal processing unit for a standardized approach and can be used for intimated processing in the proposed study, with a supported algorithm from (10). The method of detecting and validating speech signals is also proposed in this article, based on the influence of telemedicine approaches with a numerical clustering validation by (11) and (12).

The latest findings in the survey are recorded with real-time datasets as discussed in (4). This approach aims to validate treatment and handling, focusing on pandemic control and coordination. The prediction and modeling of the pandemic are discussed by Iwendi et al. (13) and Ngabod et al. (14), who propose a technique for classifying pandemic growth in smart cities. Under the process processing state, this dedicated networking model can be utilized, as discussed by Ahmed et al.,

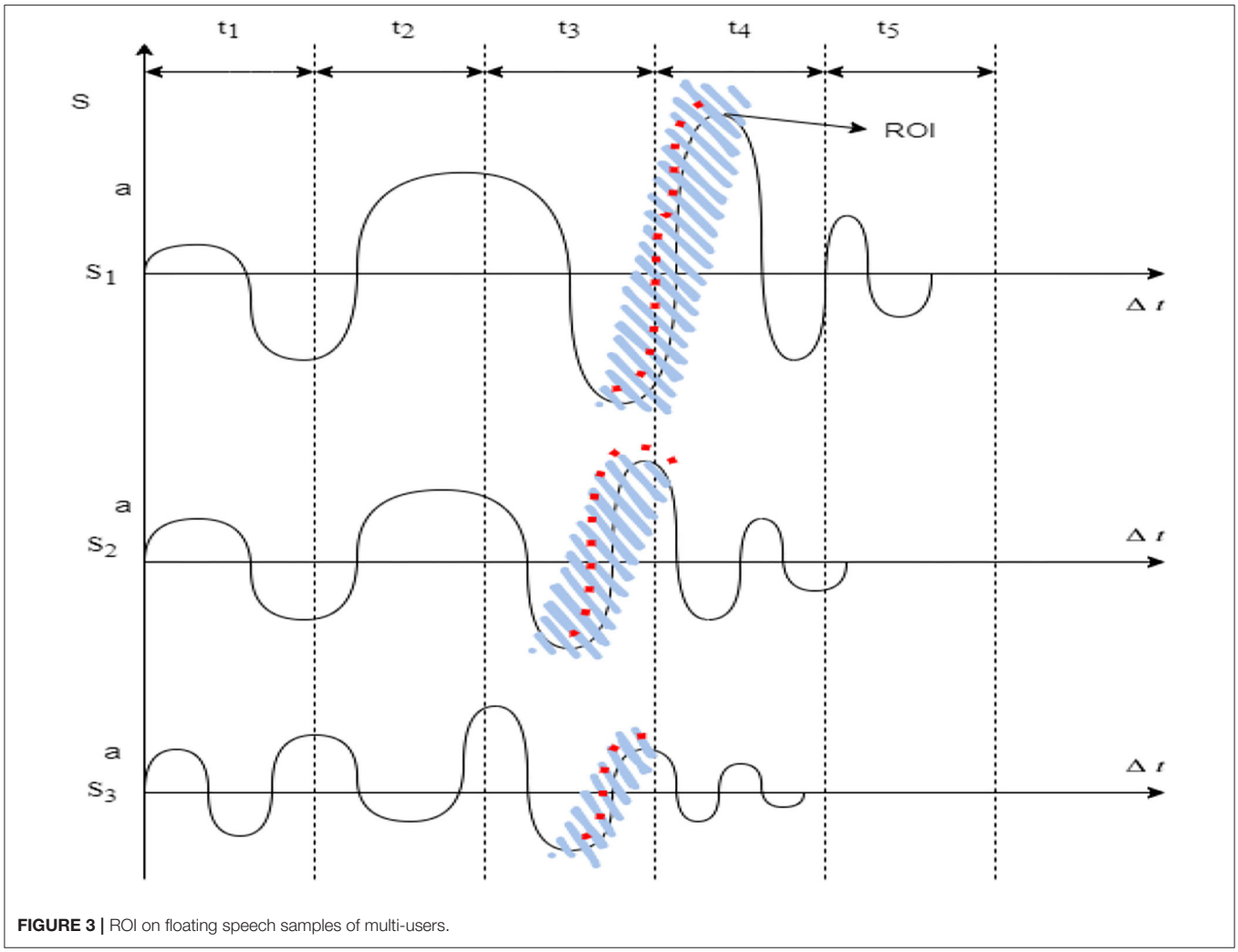


FIGURE 3 | ROI on floating speech samples of multi-users.

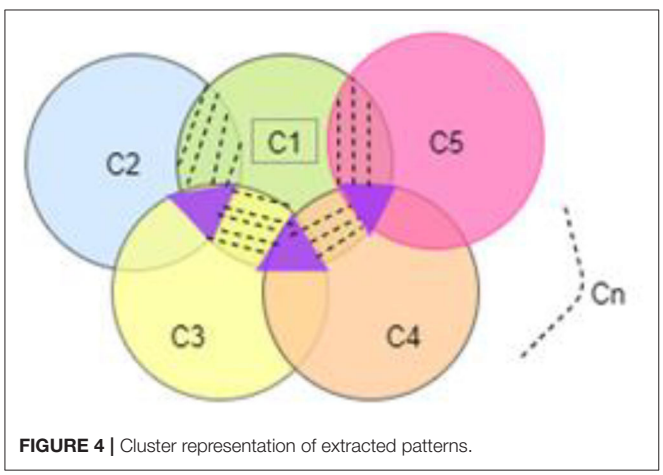


FIGURE 4 | Cluster representation of extracted patterns.

under a dynamic user cluster grouping approach (15–17). These developments have provided a reliable solution for handling pandemic data using text mining and decision support. The classification of Covid-19 studies and surveys are reported and validated by (18, 19).

METHODOLOGY

The proposed methodology aims to focus on the detection and validation of speech signals via a depression and mental disorder identification based on speech signal processing using a neural network (NN). The process is defined using mass datasets from 2,173 speech samples, as discussed in the architecture model in **Figure 1**. The agenda of the proposed technique is to restore a correlation with trained datasets in extracting and evaluating the samples of speech and classifying on demand. These speech signals are interdependent and have a higher order of distinction in recovering and validating the sample of COVID-19 patients' mental stability and sentiments (20).

The processing datasets are computed in a centralized database with user-to-user interface coordination, thereby generating a pool of databases consisting of raw and unprocessed data from the users. The process is initiated with data alignment

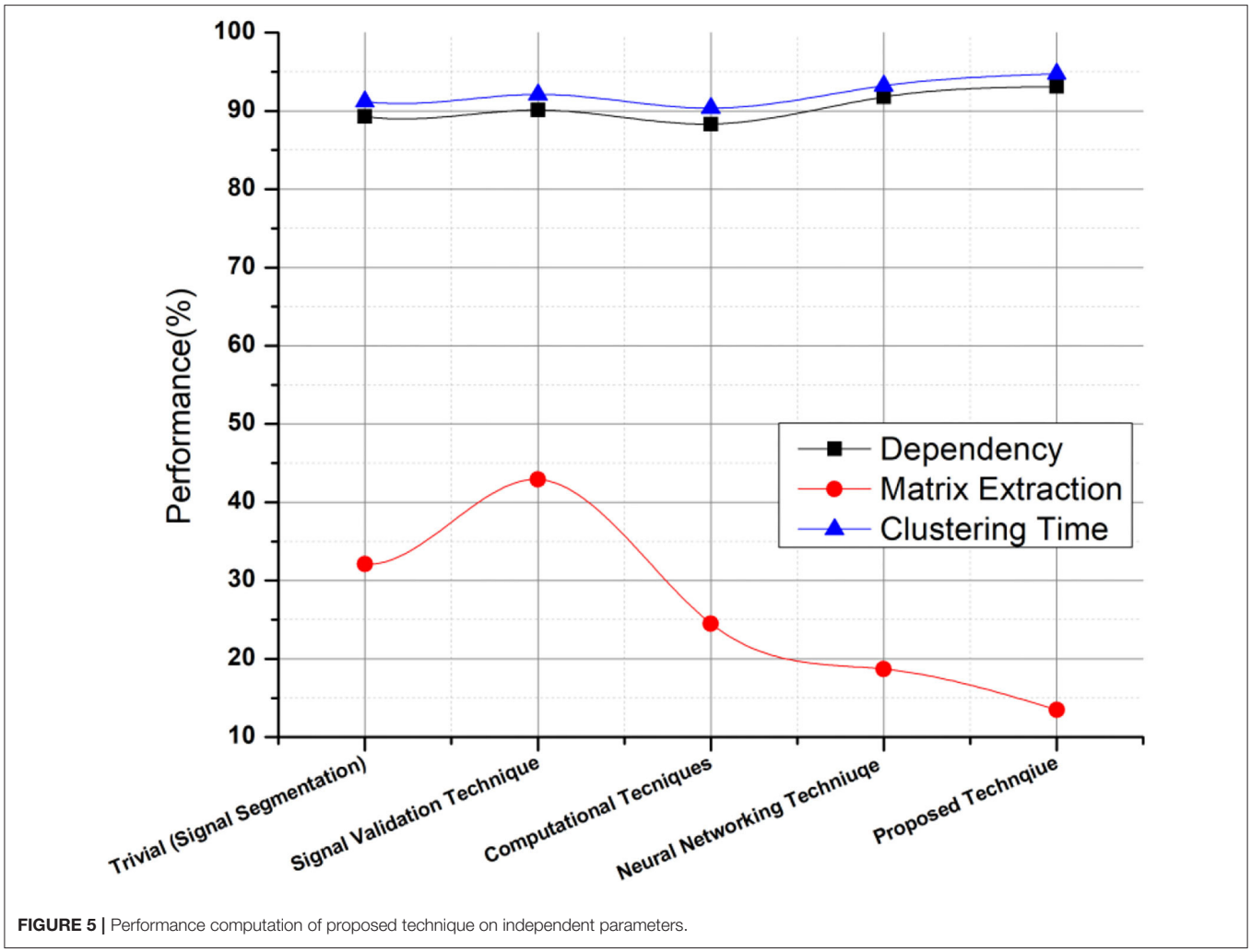


FIGURE 5 | Performance computation of proposed technique on independent parameters.

and pre-processing techniques, as discussed in the mathematical modeling of the proposed technique. The process is designed with a trained database of the speech signals with a heap address of thresholds relating to global attributes such as country, location, gender, age, and professional practice.

The trained datasets provide the threshold process for the extracted attributes of the user input signals. The process is designed with a comparative validation model to assure the process execution, as demonstrated in **Figure 2**. The comparative model evaluates the detailed execution process, such as the pattern extraction and clustering of signal samples in the form of JPEG intermediate files and a dedicated intermediate database, generating a cluster pool for segregation of samples based on ROI as demonstrated in **Figure 3**. This results in a threshold value comparison and thus provides a single decision and classification of the user's mental condition.

MATHEMATICAL REPRESENTATION

The computation of the speech signals and detection of mental stress is achieved under the processed instruction architecture, as demonstrated in **Figure 1**. The process aims to validate the signals into coordination datasets with a synchronization approach of proving learning and pooling clusters of similar patterns, as demonstrated in **Figure 3**. The mathematical approach is discussed in this section.

Attribute Extraction and Dependencies Validation

Consider a dataset *(D)* with a raw calibrated ecosystem of attributes *(A)* where each of attributes $A=\{A_1,A_2,A_3,\ldots\ldots,A_n\}$ such that each attribute (A_i) resembles the paradigm of operation, as in Equation (1).

$$A_i = \int_0^{\infty} \frac{\partial\,(A_i)}{\partial\,(D)} \tag{1}$$

Where, each of the i^{th} attributes, has a correlated paradigm of operation and process extraction. Thus, the extracted attributes (A_e) are as shown in Equation (2).

$$A_e = \int_0^{\infty} \left(\Delta D_z . \sum_{i=0}^{n} \frac{\partial A_i}{\partial D}\right) . \Delta T \tag{2}$$

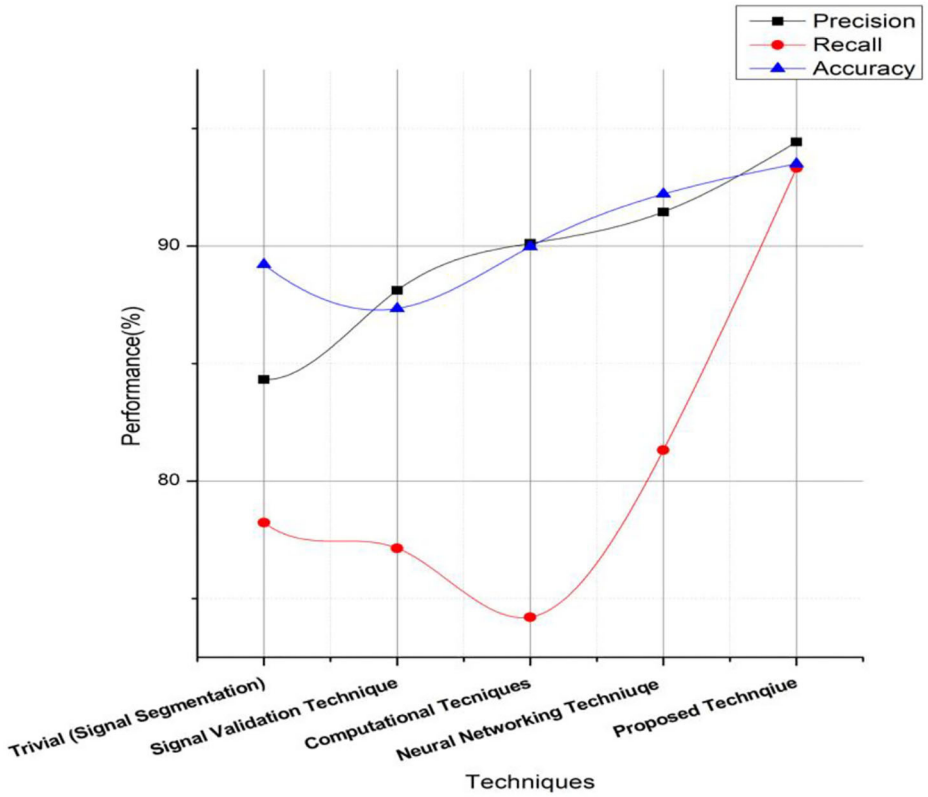


FIGURE 6 | Outcome evaluation of proposed technique.

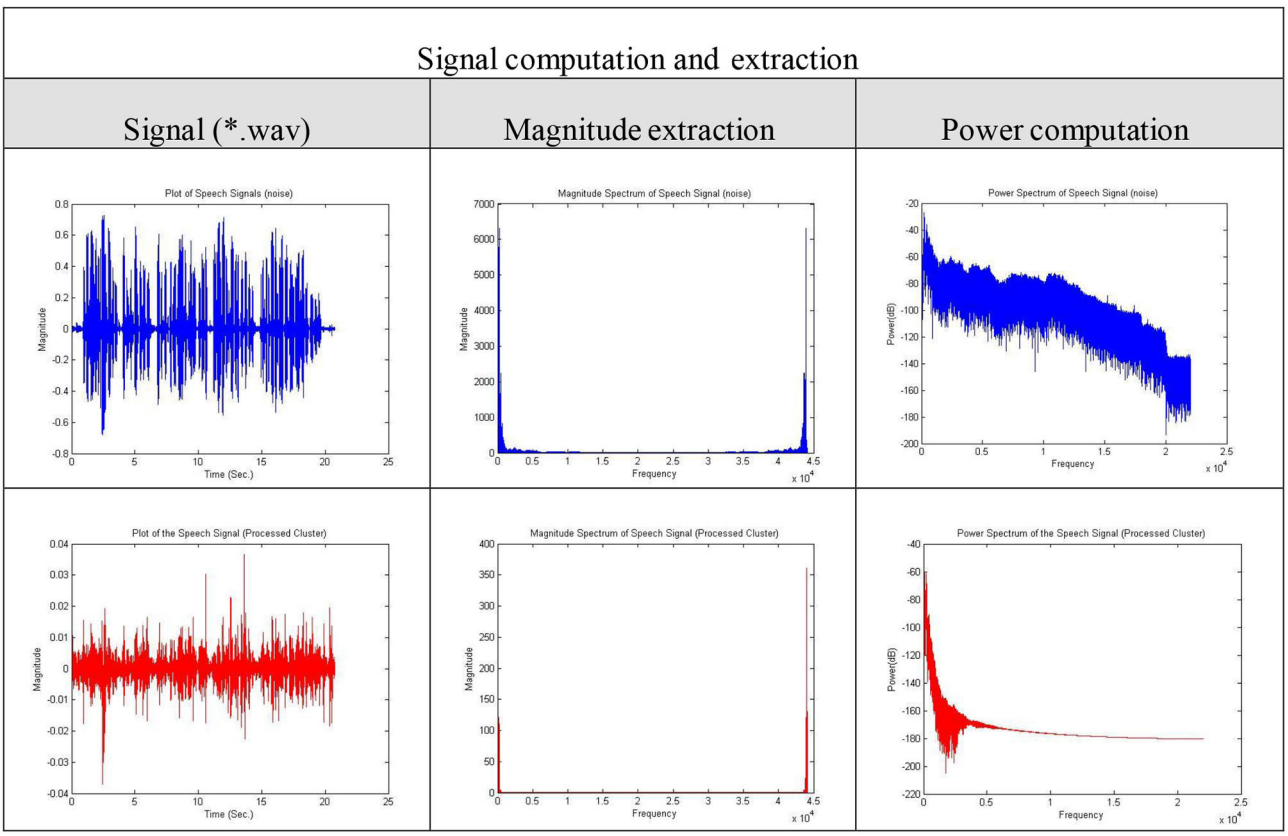
TABLE 1 | Decision support and evaluation parameters.

Patient relevance (sentiments)	Cluster range	Occurrence difference (Avg. sec)	Pattern differences	Decision accuracy (%)
COVID−19 (positive)	Average	3.211	0.327	97.23
Post COVID−19 (positive)	High	1.432	0.129	94.92
Loneliness	High	1.328	0.091	96.91
Anxiety	Average	2.114	0.181	92.17
Depression	High	0.994	0.021	97.28
Normal (Non-COVID19)	Low	5.251	0.448	95.39

TABLE 2 | Performance matrix for speech signal in mental distress validation.

Sample (Age group) (yrs)	Approach	Precision (%)	Recall (%)	Detection-score (%)	Accuracy (%)
5–9	Unclassified	92.43	89.23	89.3	90
10–15	Unclassified	95.72	93.22	93.08	95
16–25	Classified	97.12	97.48	98.1	98.48
26–35	Classified	97.78	97.53	98.9	98.97
36–55	Classified	98.3	97.12	99.12	99.23
56–60	Classified	95.33	91.23	97.12	94.19
61–75	Unclassified	96.12	89.23	91.07	88.32
76–100	Unclassified	92.31	88.7	87.2	81.2

TABLE 3 | Signal processing and analysis appendix.



Where the extracted attribute (A_e) is processed over the raw attributes set, in extracting the most relevant threshold attributes such as the peak frequency of a word or a repeated phrase of a sentence with a dilution of ΔD_z and mapping with ΔT as a threshold paradigm in validating all processing attributes (A_e) in the speech signal.

Segmentation of Samples

Samples are primarily divided into extracted attributes (A_e) sets, such that each of the attribute Region of Interest (ROI) is highlighted and marked in the entire speech signal, as shown in **Figure 3.**

Consider the segmentation (S) of the overall input signal (speech signal) with a highlighted extracted attribute (A_e). On consideration, each attribute in the signal has an occupancy time (Δt) in operating, and thus, a reflective ratio of division is processed based on CNN's evaluation paradigms.

The signal (S) of an independent sample (S_i) tends to occur in ROI in an independent location of the time matrix (Δt). Hence, the segmentation of signal (S) is as shown in Equation (4).

$$S = \frac{2\pi}{\Delta R}\int_0^{\infty}\left[\int_n^i\left(\frac{\Delta S_i}{\Delta t}\right)\right].\frac{\Delta t}{\Delta s} \tag{3}$$

$$\therefore S = \frac{2\pi}{\Delta R.\Delta S}\left(\int_0^{\infty}\left(\frac{\partial S_i}{\partial t}\right).\frac{\partial\,(A_e)}{\partial t}\right).\Delta t \tag{4}$$

Where each signal strength is measured in ΔR with a signal time Δt, for all regional attributes extraction; hence, for segmentation to be processed completely, the schematics of each attribute signal strength (ΔR) is then computed with an exhausted peak of ROI from the signal as shown in **Figure 3**.

Pattern Extraction and Schema Alignment of Datasets

The process of pattern extraction is calibrated with the internally divided segments of datasets. These datasets are processed with a frequency (f) such that the internal segments $(S)= \{S_1,S_2,S_3,\ldots\ldots,S_n\}$ has calibrated frequencies $(f)=\{f_1,f2,f_3,\ldots\ldots,f_n\}$. Thus, the inter-correlated frequencies can be defined and associated as $(S_f)=\{S_{f1},S_{f2},S_{f3},\ldots\ldots,S_{fn}\}$.

The process of pattern with speech signals is internally correlated to the amplitude of the signal (amp). Where it is represented as $f_{amp} =\{f_{amp1},f_{amp2},f_{amp3},\ldots..,f_{ampn}\}$. The amplitude of each frequency feature can be represented and extracted as shown in Equation (5).

$$S_i\,(n) = \int\left\{\sum_{k=0}^{n}\left[\alpha_i\,(k)\,.amp\,(f)\right]\right\} \tag{5}$$

Where each signal pattern (S_n) represents the overall coordination in speech signals, and the '' represents band filters of the speech signal with a coefficient of amplitude and frequency. On extraction of patterns from those correlated in Equation (5), the frequency patterns can be sorted by independent bandwidth as shown in Equation (6).

$$P_i = \frac{2\pi}{\Delta R.\Delta S}\left\{\left(\log\left(G\left(f\right)\right).\int_0^{\infty}\frac{\partial\left(S_i\left(n\right)\right)}{\partial t}\right)^2\right\} \quad (6)$$

$$P_i = \frac{2\pi}{\Delta R.\Delta S}\left\{\left(\log\left(G\left(f_{amp}\right)\right)^2.\int_0^{\infty}\frac{\partial^2\left(S_i\left(n\right)\right)}{\partial t^2}\right)\right\} \quad (7)$$

The 'P_i' on Equation 7 is the pattern of repeated learning from the CNN framework. The internal arrangements can be represented as the frequency (f) under the operation of amplitude, (mode) is represented as f_{amp}, further graded into the Gaussian constant (G). The process in Equation (7) is then concluded, as shown in Equation (8).

$$P_i = \frac{2\pi}{\Delta R.\Delta S}\left\{\left(2log\left(G\left(f_{amp}\right)\right).\int_0^{\infty}\frac{\partial^2\left(\sum_{j=0}^{n}\sum_{k=0}^{n}\alpha_i\left(j\right).amp\left(f_k\right)\right)}{\partial t^2}\right)\right\} \quad (8)$$

$$P_i = \frac{4\pi}{\Delta R.\Delta S}\left\{\left(\log G\left(f_{amp}\right).\int_0^{\infty}\frac{\partial^2\left(\sum_{j=0}^{n}\left\langle\sum_{k=0}^{n}\alpha_i\left(j\right).amp\left(f_k\right)\right\rangle\right)}{\partial t^2}\right)\right\} \quad (9)$$

$$P_i = \frac{4\pi.\log G\left(f_{\text{amp}}\right)}{\Delta R.\Delta S}.\left\{\left(\int_0^{\infty}\frac{\partial^2\left(\sum_{j=0}^{n}\left\{\sum_{k=0}^{n}\left\langle\alpha_i\left(j\right).amp\left(f_k\right)\right\rangle\right\}\right)}{\partial t^2}\right)\right\} \quad (10)$$

Thus, Equation (10) represents the coordinates of the pattern extracted and validated for the pattern with respect to a segment (S_i). Thus, on summarisation, the representation can be as $P = \{P_1,P_2,P_3,\ldots,P_n\}$ co-related to coordination of segment as $S_p = \{S_{p1},S_{p2},S_{p3},\ldots\ldots,S_{pn}\}$, where '$n$' is the last segment of given input signal.

Clustering and Classification of Datasets

Equation (10) retrieves the pattern of individual segments, and thus the coefficient of such segments are summarized and represented in $S_p = \{S_{p1},S_{p2},S_{p3},\ldots\ldots,S_{pn}\}$. Hence the clustering is shown in **Figure 4**.

The cluster (C) is retained from a group of values and its corresponding coefficients for the value re-compensation. The clusters are internally evaluated with the focus of associating.

$$C_i = \int_0^{\infty}\left\{\sum_{i=0}^{n}\left[\sum_{j=i}^{n}\left(\frac{\partial\left(P_i\right)_j}{\partial t}\right).\Delta T_i\right]\right\} \quad (11)$$

Where each cluster (C_i) is validated with a corresponding pattern coefficient and a threshold value (ΔT). The internal Threshold value (ΔT) is validated and evaluated. In summary, the clusters $(C) = \{C_1,C_2,C_3,\ldots..,C_n\}$. These clusters have an association of common patterns, for example, represented as $\{(C_i \cap C_j) \cap C_k\}$, and these associations are subjected to attribute validation, as shown in **Figure 4**.

Threshold Validation and Decision Making

The clusters and classification of speech signals using clusters are validated and approved for processing into decision making. The decision-making approach is termed with a threshold value consultation, i.e., the overall technique extracts the validated pattern coefficient and thereby synchronizes it with a relatively more and likely approach of matching and schema validation. The proposed approach typically validates the decision of signal segmentation using the threshold value toward segregating the dataset of speech signals based on emotions. These emotion-based evaluations are rather computational, and hence a most likely decision is processed.

RESULTS AND DISCUSSIONS

The proposed technique has successfully retrieved the signal attributes and the prediction ratio for evaluation. The input signals from the users via a remote connecting platform are uploaded to a centralized database in a cloud computing ecosystem using AWS-sponsored services. The datasets are processed and validated according to a multidimensional approach.

The variation of predicting the sentiments is based on the information designed and developed via clustering datasets. The prediction ratio is summarized in **Figures 5**, **6**, respectively, with a comparative evaluation from previous systems. **Table 1** shows the parameters related to the mental stress and paradigms to provide decision support. The table highlights the evaluation parameters such as the occurrence delay of a keyword in clustering, as shown in Equation (11). The supported approach thus classifies the pattern of these keyword occurrence sequences for decision making.

The results of data/signal processing and decision-making are shown in **Table 2**. The results show promising outcomes in proving a precision of 90% and higher in various users across the language and location. The results of processing a single sample are included in **Table 3**. The processing signal magnitude and the power spectrum computation demonstrate a higher order of signal clarity in analysis and validation.

CONCLUSION

The technique proposed in the present study uses neural networking terminology to learn and develop a pool of clusters and patterns to provide a systematic and reliable decision to categorize speech signals. The processing system is based on open database processing to validate the mental health conditions of users during the ongoing isolation and lockdowns caused

by the COVID-19 pandemic. The results show a promising outcome with a precision of 90% and higher accuracy across various users. The approach has a projected accuracy of 93.5% under the open validation platform on a computational evaluation. The proposed technique could be included in classifying and categorizing patients' behavior in future, with supervised approaches to keyword extraction and classification in dynamic signals.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Conceptualization and writing: SA, DS, SB, EA, and MA. Methodology: SA, DS, SB, AK, and MA. Investigation and programming: SA and DS. Resources: SB, EA, MA, and AK. Review: EA and AK. All authors contributed to the article and approved the submitted version.

FUNDING

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP-2021/164), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

1. Ahmed ST, Basha SM, Arumugam SR, Kodabagi MM. *Pattern Recognition: An Introduction*. MileStone Research Publications (2021).
2. Lin C, Ibeke E, Wyner A, Guerin F. Sentiment–topic modeling in text mining. *Wiley Interdisc Rev.* (2015) 5:246–54. doi: 10.1002/widm.1161
3. Singh S, Roy MD, Sinha CPT, Parveen CPTMS, Joshi CPTG. Impact of COVID-19 and lockdown on mental health of children and adolescents: a narrative review with recommendations. *Psychiatry Res.* (2020) 2020:113429. doi: 10.1016/j.psychres.2020.113429
4. Naik PA, Owolabi KM, Zu J, Naik MUD. Modeling the transmission dynamics of COVID-19 pandemic in caputo type fractional derivative. *J Multisc Model.* (2021) 20:1–20. doi: 10.1142/S1756973721500062
5. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. *N Engl J Med.* (2020) 383:510–2. doi: 10.1056/NEJMp2008017
6. Spoorthy MS, Pratapa SK, Mahant S. Mental health problems faced by healthcare workers due to the COVID-19 pandemic–A review. *Asian J Psychiatry.* (2020) 51:102119. doi: 10.1016/j.ajp.2020.102119
7. Cullen W, Gulati G, Kelly BD. Mental health in the Covid-19 pandemic. *QJM Int J Med.* (2020) 113:311–2. doi: 10.1093/qjmed/hcaa110
8. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multi Syst.* (2021) 1–15. doi: 10.1007/s00530-021-00774-w
9. Ahmed ST, Sankar S, Sandhya M. Multi-objective optimal medical data informatics standardization and processing technique for telemedicine via machine learning approach. *J Amb Intell Hum Comput.* (2020) 12:5349–58. doi: 10.1007/s12652-020-02016-9
10. Ahmed ST, Sandhya M, Sankar S. An optimized RTSRV machine learning algorithm for biomedical signal transmission and regeneration for telemedicine environment. *Proc Comp Sci.* (2019) 152:140–9. doi: 10.1016/j.procs.2019.05.036
11. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in Pakistan. *Comput Mater Continua.* (2021) 69:1253–69. doi: 10.32604/cmc.2021.015655
12. Kumar SS, Ahmed ST, Vigneshwaran P, Sandeep H, Singh HM. Two phase cluster validation approach towards measuring cluster quality in unstructured and structured numerical datasets. *J Ambient Intell Hum Comput.* (2020) 12:7581–594. doi: 10.1007/s12652-020-02487-w
13. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. Jo O. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
14. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng.* (2021) 18:8444–61. doi: 10.3934/mbe.2021418
15. Ahmed ST, Sandhya M, Sankar S. TelMED: dynamic user clustering resource allocation technique for MooM datasets under optimizing telemedicine network. *Wireless Pers Commun.* (2020) 112:1061–77. doi: 10.1007/s11277-020-07091-x
16. Al-Shammari NK, Syed TH, Syed MB. An Edge–IoT framework and prototype based on blockchain for smart healthcare applications. *Eng Technol Appl Sci Res.* (2021) 11:7326–31. doi: 10.48084/etasr.4245
17. Ahmed ST and Sankar S. Investigative protocol design of layer optimized image compression in telemedicine environment. *Proc Comput Sci.* (2020) 167:2617–22. doi: 10.1016/j.procs.2020.03.323
18. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc.* (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
19. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021) doi: 10.1049/IPR2.12222
20. Reddy PK, Reddy TS, Balakrishnan S, Basha SM, Poluru RK. Heart disease prediction using machine learning algorithm. *Int J Innov Technol Expl Eng.* (2019) 8:2603–6. doi: 10.35940/ijitee.J9340.0881019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ahmed, Singh, Basha, Abouel Nasr, Kamrani and Aboudaif. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fpubh.2021.798905

Ontology-Enabled Emotional Sentiment Analysis on COVID-19 Pandemic-Related Twitter Streams

Senthil Kumar Narayanasamy[1], *Kathiravan Srinivasan*[2], *Saeed Mian Qaisar*[3] *and Chuan-Yu Chang*[4,5]*

[1] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [2] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, [3] Electrical and Computer Engineering Department, Effat University, Jeddah, Saudi Arabia, [4] Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, [5] Service Systems Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

The exponential growth of social media users has changed the dynamics of retrieving the potential information from user-generated content and transformed the paradigm of information-retrieval mechanism with the novel developments on the concept of "web of data". In this regard, our proposed Ontology-Based Sentiment Analysis provides two novel approaches: First, the emotion extraction on tweets related to COVID-19 is carried out by a well-formed taxonomy that comprises possible emotional concepts with fine-grained properties and polarized values. Second, the potential entities present in the tweet can be analyzed for semantic associativity. The extraction of emotions can be performed in two cases: (i) words directly associated with the emotional concepts present in the taxonomy and (ii) words indirectly present in the emotional concepts. Though the latter case is very challenging in processing the tweets to find the hidden patterns and extract the meaningful facts associated with it, our proposed work is able to extract and detect almost 81% of true positives and considerably able to detect the false negatives. Finally, the proposed approach's superior performance is witnessed from its comparison with other peer-level approaches.

Keywords: sentiment analysis, emotion ontology, natural language processing, twitter streams, latent Dirichlet allocation, SPARQL

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Muzammil Khan,
University of Swat, Pakistan
Muhammad Adnan Zahid Chudhery,
University of Science and Technology of China, China
Ebuka Ibeke,
Robert Gordon University, United Kingdom

***Correspondence:**
Chuan-Yu Chang
chuanyu@yuntech.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *20 October 2021*
Accepted: *04 November 2021*
Published: *06 December 2021*

Citation:
Narayanasamy SK, Srinivasan K, Mian Qaisar S and Chang C-Y (2021) Ontology-Enabled Emotional Sentiment Analysis on COVID-19 Pandemic-Related Twitter Streams. Front. Public Health 9:798905. doi: 10.3389/fpubh.2021.798905

INTRODUCTION

Emotion Analysis helps us to understand and to retrieve the potential emotional information from any user-generated content. The user-generated content can be text, images, animations, videos, scan images, etc. But the text is predominantly used due to the emergence of social media platforms and other sources as well. Extracting emotions from the text document is a challenging task as it covers a wide range of ambiguities and anomalies that persist over the text content. Emotion Analysis has been categorized into different models, such as the corpus-based model, appraisal-based model, and knowledge-based model. The lexical affinity of a word or a phrase is determined based on the probability of emotion attached to each term and the cognitive factor will be inclusively considered for evaluation. For example, the term, "eager" takes a high probability score against "curious" than "willing." To shun these difficulties, an ontology-based analysis has been used predominantly to disambiguate the terms and to yield a high precision score for the emotional terms.

At first, the term, ontology was inhibited from psychology but it started making its footprint in computer science from the year 1980s. Ontology-based emotion extraction on text documents has been a research focus for many applications and it has gained huge momentum in recent years (1–3). As human beings always use natural language to represent the domain of the specific text document, the ontologies have been using formal language representations to describe the domains of the input document. The use of ontologies has been widely applied in many research areas, such as artificial intelligence, entity extraction, Semantic Web, collaborative software development, and many more (4, 5). Ontologies provide huge benefits, such as conceptualization, reusability, sharing of the resources, and coreferencing the terms. It is indeed a cost-effective process for dealing with text analysis and is very robust for many application areas as mentioned above. Many research works have been implemented with ontologies to remove the ambiguity that persists in the textual content. A structure that provides a formal description of a standard representation of real-world concepts can improve the understanding of these concepts. COVID-19 is the disease caused by SARS-CoV-2, the coronavirus that emerged in December 2019 (6–11). The lack of existing studies allows us to explore this ontology-based emotion extraction on COVID-19 datasets.

The conventional Sentiment Analysis is normally applied to the text to determine whether the given text expresses a positive sentiment or a negative sentiment. Generally, the Sentiment Analysis has been carried out for knowing the feedback or the opinion of the customers about the products they had purchased. In particular, the following two Sentiment Analysis approaches were usually performed to get the sentiment score of the text documents: lexicon-based approach and machine learning-based approach. The first approach segments the text into appropriate morphological lexicons and then extracts the opinion words which are usually expressed either positive or negative. A dedicated dictionary that has been established for opinion words is called a lexicon dictionary and mapping of the opinion words on the lexicon-based approach has been done based on this lexicon dictionary. But in this case, tweets are not considered because of the underlying fact that tweets are within 140/280 characters and normally, words in a tweet are condensed into canonical forms due to their limited length. The machine learning-based approach has now gained huge momentum for many research activities and in that case, a specialized sentiment classifier has been trained for the textual corpus to predict the polarity of the text. Typically, the sentiment classifier has trained the model based on bigram or n-gram textual representation and classifies the text into the respective sentiment scores. The major drawback of this approach is the manual labeling of the training datasets and it has not been worked out well for tweets or any short messages.

The ontology-based approach proposed by Ali et al. (12) indicates that the sentiment score has not been made possible based on the emotional polarity but on the accuracy score obtained from the ontology. The ontology-based model helps to extract the emotions from the tweets based on the factors, such as concepts, the relationship between the concepts, characteristics of individual concepts, and external source document support for disambiguation. In this connection, Semantic Web technologies have been used to construct the ontology for extracting the emotions from the text documents and allow for sharing and reusing of the potential data for various applications. Resource description framework (RDF) is used to identify the resources by Uniform Resource Identifier (URI), and Resource description framework schema (RDFS) helps to organize and formulate the content in the machine-understandable format. The SPARQL query is used to fetch the disambiguous results from the RDFS graph. Therefore, we proposed the novel approach to segment the tweets into appropriate morphological textual representations and to train the classifier to distinguish the polarity of tweets using Semantic Web technologies and natural language processing (NLP) toolkits.

The primary objective of this research work is to quantify the efficient classification of the extracted emotional entities from tweets using the proposed emotional ontology. Following are the outlined objectives of the proposed work:

- The domain entities have been classified with appropriate upper ontology classes and it measures the correctness concerning the metrics given in the Emotion Ontology.
- Selecting and quantifying some of the general domain emotional entities from the tweets that have significant relevance in the upper ontological class of Emotion Ontology.
- Determining the word embedding and taxonomy-based similarity measures for the extracted emotional entities.
- Generating relevant descriptions for those extracted emotional entities and assigning the relevance score based on the results obtained from the SPARQL query.

The rest of the paper is organized as follows: Section Related Works summarizes the existing works based on semantic similarity measures, ontological features, emotion analysis, and sentiment analysis of social media content. Section Proposed Approach for Ontology Creation highlights the important features of Domain Ontology and semantic enrichment of entity categorization. Section Utilization of Emotion Ontology describes the core aspects of designing the semantic-based emotion ontology and taxonomy-based entity classification and disambiguation. Besides, it also explains the importance of ontology modularization for partitioning the large-scale ontologies into some self-contained modules. Section Conclusion delineates the polarity calculation and determines the performance of emotion ontology for the fine-grained measures.

RELATED WORKS

Extracting emotions from the documents is a tedious task as it involves a wide range of subjects, such as psychology, anthropology, society, and biology. In psychology, emotion is defined as the study of different orientations of human beings and the theories of emotions have been represented using cognitive psychology. Even though many cognitive models had been evolved to denote the emotions incurred in the documents, the most widely implemented affective computing models are dimensional (13), categorical, (14), and appraisal (15). Lang (13)

has extensively studied and analyzed the expressions of emotions for effective detection and categorization. He had classified the emotions into three vital categories, namely, subjective emotions, behavioral emotions, and psychophysiological emotions. The subjective emotions were classified based on perceived emotions vented by the user. The behavioral emotions were distinguished by facial, gestural, and speech paralinguistic attributes. The psychophysiological emotions were identified by the heart rate, electroencephalographic results, galvanic skin values, etc.

Later, Lacy (16) had delineated the behavioral models with emotional processes and the methods required to indicate the various stages of emotions are represented by cognitive psychology. They further analyzed the emotional processes very deeply and came out with three critical parameters, to access the behavioral model, such as emotion itself, emotional context, and multimodal behaviors. The study added that time-of-event is a critical factor for representing the emotional context. On the contrary, Baldauf et al. (17) had proposed a novel model in which they assessed the behavioral model in terms of location, time, the person involved, the state of the person at the time of the event, societal activity, and instruments or devices used. These parameters were used for emotional context and augmented the emotional credibility for research analysis.

Later et al. (18) have extended the sentiment analysis of Twitter streams by computational frameworks and analyzed the emotions expressed in Twitter streams by certain questionnaires. The answers pertaining to the questionnaires were later analyzed by the popular probabilistic topic model, latent Dirichlet allocation (LDA) algorithms that effectively disambiguated the ambiguous emotions present in the tweets. Their work highlighted the various factors of emotion discovery shifted among various users' conversations on Twitter.

The emergency of social media has created a huge opportunity for social users to interact and discuss a wide range of topics prevalently happening around the world. These discussions paved the way for analyzing the significant interest of the social media users over the topics and performing sentiment analysis, such as emotion detection and recent opinion mining. According to Ren and Hong (19), emotion analysis had been widely implied in various domains, such as e-mail content, novels, online news content, blogs, dashboards, and other social media content. Researchers, such as Oliveira et al. (20) and Ren et al. (21) have studied the consumers' views on many online products and their purchases and later predict the growth of stock markets. They forecasted the growth of stock markets based on discussions that happen on Twitter and followed the lexical heuristic approach to filter out the basic emotional values presented in the tweets.

The researchers, Baldoni et al. (22) and Dey et al. (23) have implemented the emotion extraction process first by n-ary relations. The complex relations were factorized into some set of binary relations and the potential named entities were extracted from the text. Extracting potential named entities from the text is a complex task and it requires a well-equipped classifier to identify and distinguish the potential entities from the text. So, the classifier has been well trained to suit the needs of the entity extraction and in this case, the learning approach techniques have been used and in particular, the maximum entropy model has been followed for entity extraction. Once the potential named entities have been identified, the complex relationship between the entities has been reconstructed using a simple directed graph. The three significant approaches followed before the emotional analysis processes include (i) rule-based approach, (ii) learning approach (maximum entropy), and (iii) graph-based approach.

Earlier, open information extraction (OIE) has gained huge momentum in extracting the entities and relation sets from the documents. OIE is a full-fledged extraction framework Daniel et al. (24) for both structured and semi-structured documents. OIE enables the extraction process in place of DBpedia and LinkedGeo data for effective entity detection and categorization. Further, it paves way for entity triple formation using the Semantic Web technologies, such as RDF/RDFS and Web Ontology Language (OWL).

Most of the semantic similarity measures attempt to emulate the human ability to evaluate the level of relativeness between the words according to their semantic evidence. According to Madani et al. (25), semantic measures assess the quality of the semantic associations between the words, as indicated by the investigation of semantic proxies (words, senses). For instance, a semantic similarity measure would not think about the two ideas "sloth" and "monkey" to be comparative, regardless of whether the vast majority think sloths to be monkeys. Given that semantic estimates target contrasting things concurring with their significance caught from semantic proof Mozafari et al. (26), it is hard to additionally characterize the thought of semantic measures without characterizing the ideas of meaning and semantics. Likewise, the two ideas, "tea" and "cup" are in this way exceptionally related despite the way that they are not comparable: the idea, "tea" alludes to a drink, and the idea, "cup" alludes to a vessel. In this manner, the two ideas share not many of their constitutive properties. This features a potential understanding of the idea of semantic similarity, which can be comprehended in terms of replacement, i.e., assessing the suggestion to substitute the analyzed components: Tea by Coffee or Tea by Cup. In fact, word-to-word semantic similarity is sometimes assessed not just considering (close) synonymy, or the lexical relations which can be considered as comparable to the ordered connections for words, e.g., hyponymy and hypernym.

Table 1 highlights the performances laid out on semantic similarity measures of different domains.

PROPOSED APPROACH FOR ONTOLOGY CREATION

At first, the proposed approach deals with the creation of Domain Ontology for the bio-medical-related subjects to provide an accurate sentiment score for the opinion words present in the COVID-19 tweets. Here, the use of Semantic Web technologies gains a tremendous prospect to create Domain Ontology for the specific subjects, and further, the opinion words from the subject have been enlisted for the ontology with appropriate properties and literals. The Domain Ontology plays a seminal role in supplementing the feature extraction process to yield the precise sentiment score for the COVID-19-related tweets and

TABLE 1 | Highlights of semantic similarity measures and dataset assignments.

S.No	Model references	Year	Core evaluation method	Data size	Inter-agreement
1	Dridi and Recupero (27)	(2019)	Semantic similarity Association between the pairs of nouns	40	$\Upsilon = 0.784$
2	Mozafari and Tahayori (26)	(2019)	Semantic relatedness exists among the pairs of nouns.	Two sets 150/200	$\rho = 0.661$
3	Dragoni et al. (28)	(2018)	Semantic proximity of word pairs on the basis of synonymy questions	50	NA
4	Chen et al. (29)	(2017)	Semantic similarity evaluation of words based on analogy questions	390	NA
5	Li et al. (30)	(2017)	Semantic Associativity between 25 medical words	34	$\Upsilon = 0.51$
6	Zhu and Iglesias (31)	(2016)	Semantic similarity and proximity score of pairs of UMLS concepts (domain: medical)	571	$r = 0.50$ (sim) $r = 0.47$ (rel)
7	Bruni et al. (32)	(2014)	Semantic relatedness of pairs of words	2500	$\rho = 0.74$
8	Hill et al. (33)		Semantic similarity of pairs of words	999	$\rho = 0.7$

increase the accuracy rate of the results. The proposed system is divided into two crucial phases. (i) Creation of Domain Ontology for the selective subject and (ii) feature extraction of opinion words included in the Domain Ontology. **Figure 1** illustrates the proposed architecture for emotions extracted from the tweets.

Creating the Domain Ontology

Normally, the Domain Ontology can be created either by utilizing the existing off-the-shelf ontologies or developing the ontologies based on the suitable requirements. There were many off-the-shelf ontology resources available for specific domains in the Semantic Web Forums that paved a way for the researchers to utilize these resources to map the properties of the ontology and opinion words (34). Many software solutions have been completely dedicated for knowledge-based entity extraction and deployed further with the extraction of context-specific ontology/structured terminologies, such as the disease ontology (DO), unified medical language system (UMLS), WordNet, Medical Subject Headings (MeSH), Gene Ontology, etc. To support the OWL, the standard RDF grammar has been designed with stipulated functionality (31). These ontological frameworks (28) can be used to define the terms and entities generically. In **Table 2**, we have given the name of the tools used for ontological design, supporting formats, libraries, command-line interface (CLI), and source code library (LIB). As the ontological frameworks support the programming languages, such as JAVA and PYTHON, the semantic similarity measure can be taken into consideration for some pair of concepts (P) or some association between the groups of concepts (G).

Formal Concept Analysis for Ontology Creation

Formal concept analysis (FCA) is a mathematical model that has been mostly applied in knowledge representation and information extraction. The sole objective of FCA is to pave the way for creating the Domain Ontology by a step-by-step approach and by considering all the ambiguities that persist on the properties of the Domain Ontology. With the advent of recent Semantic Web Technologies, the OWL helps to derive the potential objects and their properties and establish the Domain Ontology for accurate knowledge representation.

Basic Elements of FCA in Domain Ontology

The fundamental building block in FCA is the concept that can be derived from two valuable sets: Extension and Intention. The Extension is a set of objects collected for the Domain Ontology in considering the potential opinion words listed for the specific subjects. The set of objects has been chosen based on some grounding rules formulated for creating the ontology. The Intention is a set of attributes for every selected object and each attribute has some inherent properties and literals to represent in the hierarchy of domain levels. Every object that represents the concept has its attributes in the intention and likewise, each attribute linked to the concept of the Domain Ontology can be further shared with all the other objects of the Extension. This mutual connection that exists between the objects and attributes is described through the mathematical representation called Formal Context. A Formal Context is a triple that can be represented as K (O, A, I). Here O is a collection of objects for the Domain Ontology, A is the collection of potential attributes of the objects, and I is a binary correlation that exists between the objects and attributes; $I \subseteq O \times A$, where (o, a) $\in$ I (i.e., "object o has to attribute a").

Method of Ontology Learning

Ontology Learning or ontology acquisition is the inbuilt process of automatically inheriting the concepts and their associated properties from the other ontology resources or datasets (1). This capability has helped the researchers to save time and customize their Domain Ontology as per their requirements. In this study, we have used OntoGen, a semi-automatic ontology editor that reduces the overall deploying time of the ontology, as well as the complexity, which persists over the hierarchy of the ontology. In nutshell, this editor helps resolve the conflict between the ontology editors and domain experts because both do not hold the required ontology-engineering skills. That is, OntoGen is an

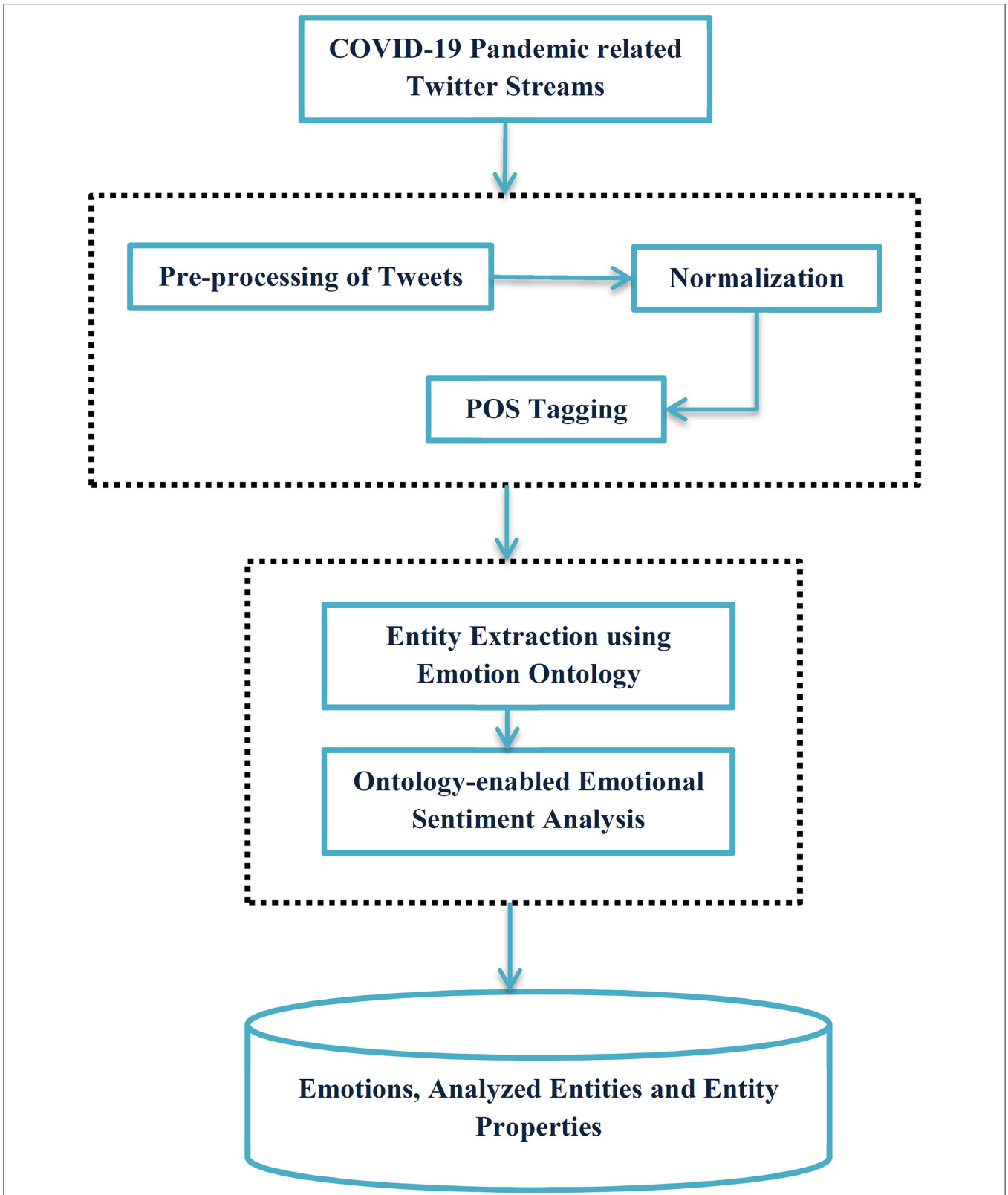


FIGURE 1 | Proposed architecture for emotion extraction from Tweets.

TABLE 2 | Semantic tools and ontology design for CLI and LIB.

Name	Ontology	Types	Measures	Language
SML	OWL,RDF, and OBO	CLI, LIB	P, G	Java
FastSemSim	OBO and others	CLI, LIB	P	Python
OntoSim	OWL, RDF	LIB	P	Java
YtexSemanticSimilarity	SEE DOC	LIB	P, G	Java
SimilarityLibrary	Wordnet, MeSH, GO	CLI, LIB	P, G	Java / Python
OWLSim	OWL, RDF, OBO	LIB	P	Java / Python

interactive editor that can help the domain experts to accept or reject the concepts or objects, or properties of the ontology, based on their requirements, and suggests the appropriate concepts and relations to the ontology. The manual adjustments of assigning instances to the concepts can be done at any level of the ontology hierarchy and thereby reduce the complexity of the overall ontology representation.

Semantics Entailment

The domain ontology has been created *via* FCA and Ontology Learning and it enriched the taxonomy of a set of concepts and its potential attributes. To boast the semantic representation of the concepts, the ontology has now been filled with appropriate synonyms and hyponyms of the corresponding attributes. For instance, the term, "apple" comes under the class, "fruit" as well as "phone"; further, their hierarchy extends above to "vegetables" for "fruit" and "electronics" for "phone." The synonyms and hyponyms can be fetched through the well-formed lexical corpus called WordNet (35, 36). With the assistance of WordNet, we extracted the appropriate synsets for the word and appended the same to the ontology and associated with the appropriate attributes of the object. We have been using the Semantic Web Language for the ontology called OWL which has the following three dedicated categories: OWL DL, OWL Lite, and OWL Full (16). To augment the synsets extracted from WordNet to the Domain Ontology, we preferred to use OWL DL which has the syntax owl: subPropertyOf and owlequivalentProperty. Through these OWL properties, we updated the synonyms and hyponyms, respectively, into the ontology.

Sentiment Analysis on Twitter Streams

After the creation of domain ontology, the next phase of the operation largely relies on performing the Sentiment Analysis of the tweets and distinguishing the polarity of the extracted opinion words from the tweets. The whole process constitutes of filtering the set of tweets from the COVID-19 datasets and then extracts the potential opinion words from the tweets and maps the opinion words in the domain ontology for getting the sentiment score (37, 38). To obtain the sentiment score of the opinion words, we have scaled up three distinct procedures: (i) querying the Domain Ontology for retrieving the correct attributes of the object, (ii) distinguishing the ambiguities that persist over two or more objects, and (iii) disambiguating the objects with necessary attributes and literals.

Step#1: Utilization of the Ontology

The Domain Ontology has been used to extract the appropriate attributes for the opinion words and to fetch equivalent classes for the objects with necessary attribute references. This task can be performed through Jena Fuseki, which is a Java API for handling RDF/s and OWL codes (39). Generally, JENA API is used to filter the results in the form of a triplet which is represented as Subject, Predicate, and Object (SPO). This triplet is used to discriminate the classes of the ambiguous attributes and return the results either in the form of an RDF Graph or as JSON format. Once we obtain this sort of ontology-based hierarchy model for classes and attributes, retrieving the triplet would be easier and this, in turn, results in matching the opinion words in the Domain Ontology.

Step #2: Identify the Ambiguous Entities in the Tweet

In the preprocessing steps of tweet normalization, we removed all the special symbols like @, #, and other uniform resource locators (URLs), and then considered the words which are deemed to be opinion words after verifying them with the opinion-lexical database. The pre-processed tweets are then added to the second phase of the Sentiment Analysis and here we matched the opinion words which are also called entities in the Domain Ontology. The mutual mappings between the opinion words in tweets and their equivalent attributes in the Domain Ontology pose an ambiguity and create a separate list of ambiguous entities for further disambiguation. For logics and inferences possessed in the Ontology, we have used the first-order logics as well as description logics to express the relations given in the Emotion Ontology.

Step #3: Disambiguate the Objects

The ambiguous list of entities is then processed for entity disambiguation which is the crucial task of this research work. The core idea behind the disambiguation process is that the mapping between two objects with dissimilar attributes in the Domain Ontology is identified first and then a comparison of properties and literals of two objects can be measured using semantic similarity-based algorithms i.e., Word2Vec and disambiguate the entities which hold the highest similarity score. The fundamental principle of ontology lies in disambiguation so that the original sense of the attributes can be obtained only at the leaf nodes of the hierarchy and hence the semantic similarity measures, such as Word2Vec have been selected to delve deep into the hidden layers of the neural network to get the precise results for the target words. Algorithm 1 illustrates

Algorithm: 1 Emotional Ontology-Based Entity Extraction and Weight Calculation.

Input: Parse Tweets and Extract Named Entities $NE = (e_1, e_2, \ldots . e_n)$
Output: Collection of entities mapping with emotion ontology
Begin
  For every entity e_i *in NE do*
    For every domain D_j *in EmoOnto do*
    if the term e_i *present in EmoOnto then*
    Add e_i *into ResultSet R*
    end if
For every entity $e_i \in$ *ResultSet R do*
        Calculate the semantic weight sw of term e_i *as,*
        $wt = [\Sigma$ *(Count of* e_i *in* $d_j)$ *]* * *[log(n/df)]*
End for
End for
End for
End

the Emotional Ontology-Based Entity Extraction and Weight Calculation process.

UTILIZATION OF EMOTION ONTOLOGY

According to Li et al. (40), ontologies are originally stated as a formal, explicit specification of a shared conceptualization. It originally deals with knowledge representation in the hierarchy of concepts ordered in the ontology and it shared the common attributes among other classes in the hierarchy to represent the types, relations, properties, and other interrelation concepts that lie in the ontology. Ontology plays a crucial role in the knowledge extraction process and represents some conventional extraction models, such as keyword-based extraction, bag-of-word models, and other statistical analysis that failed to yield accurate results. Ontologies help share the common attributes and they are made machine-readable. Ontologies play a pivotal role in establishing the mutual connection between class hierarchies and other concepts underlying the objects. Further, the ontologies are very useful for exact labeling and categorizing the mutual relationship that exists between the objects and their properties. The ontologies are defined in the following four forms: (1) an entity that links the object, (2) the relation between the entity, (3) object relationship, and (4) properties that connect the objects. The use of ontology is now gaining momentum and the following are some potential reasons for building the ontology for the given requirements: Examining domain-specific knowledge, distinguishing the domain suppositions, reusing the domain knowledge for diverse applications, converging the domain knowledge and functional designs, and sharing the knowledge with other software bots.

Emotion Ontology Corpus Creation

Eight emotions were selected based on the recommendation given by Ekman (14) and observed that these eight emotions have been widely used in the COVID-19 datasets. We have populated these eight emotions into the Emotion Ontology and represented them as OWL Class (41). A dedicated list of emotional categories of terms has also been chosen based on the psychological models (42) and we structured them in the Emotion Ontology as given in **Figure 2**. Further, 14 subsets of emotions are added as subclasses for the above eight emotions. Besides, the emotions can be further classified into four types: *Basic Emotions* (joy, confusion, anger, disgust, fear, sadness, love, and surprise), *Mid Emotions* (distraction, boredom, acceptance, apprehension, interest, serenity, pensiveness, and annoyance), *Intense Emotions* (ecstasy, amazement, vigilance, grief, admiration, loathing, rage, and terror) and *Complex Emotions* (disproval, love, submission, optimism, awe, contempt, aggressiveness, and remorse). The Emotion Ontology helps to reduce the complexity of the redundant emotion population and also paved the way for detecting the emotions represented in the tweets. Besides, Ontology minimizes the non-emotional terms, which would be very large in Twitter Streams.

The Domain Ontology has been created manually with the help of a psychological-emotional words list and WordNet. But when it comes to the reusability of several classes belonging to different domains, the manual population of ontology would be a hard task. Further, it would lead to duplication of many more properties and their values. Therefore, effective utilization of Semantic Web technologies, such as RDF/RDFS, SPARQL, and OWL would primarily facilitate the mutual connection that exists over other domain ontologies, such as the one given in the Linked Open Data (43). In this proposed approach, Emotion Ontology has been created based on the effective reuse of existing ontologies and with well-formed semantic vocabularies, such as a friend of a friend (FOAF), semantically interlinked online communities (SIOC), Dublin core, etc. The SPARQL query has been generated to map the semantic vocabularies and other domain ontologies appropriately. **Figure 3** illustrates the generation of Emotion Ontology.

Retrieving and Annotating the Twitter Streams

In recent times, machine learning techniques have been used to automatically annotate the text documents and to train the classifiers according to the needs of the ontological requirements and domain hierarchy (2, 4, 29). But the machine learning techniques are not suitable for unstructured documents and they further lead to many serious implications, such as morphological change, lexical error, sense overlap, and ambiguous annotation for entities. Hence, we proposed a novel approach called ontology-based entity extraction and constructed a novel Emotion Ontology. As the Twitter Streams contain diverse information from different sources, the distribution of emotions that persisted on these tweets is deemed essential for extracting more emotions which further helps in tracking the implicit emotions of individual users on Twitter.

To gain a high recall score for the emotion extraction from the tweets, we have also crawled for slang words and out of vocabulary (OOV) words present in the tweet. The slang words and OOV words were disambiguated and we generated the correct term for the slang words using online slang dictionaries,

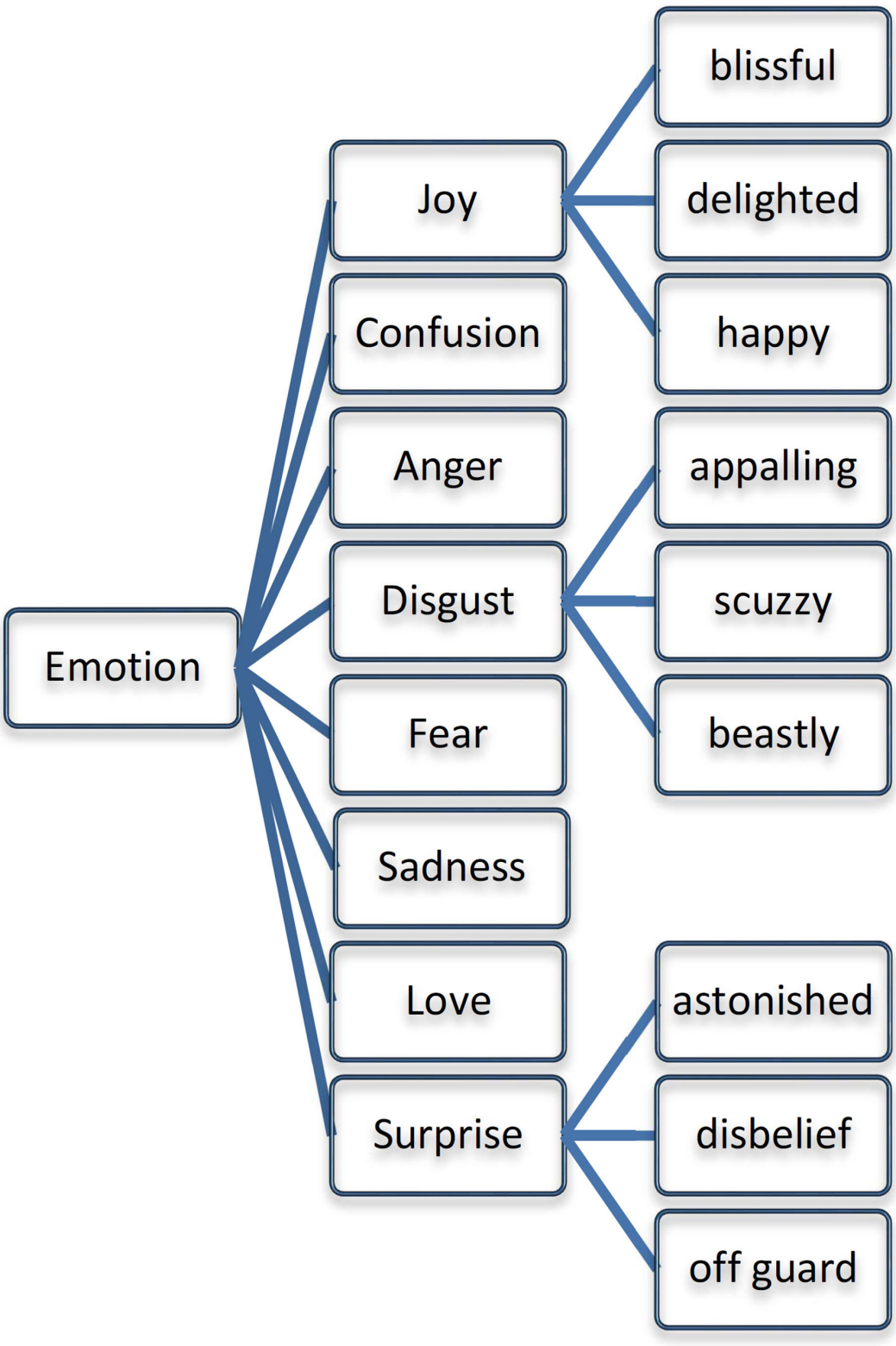


FIGURE 2 | Fundamental emotions list for ontology population.

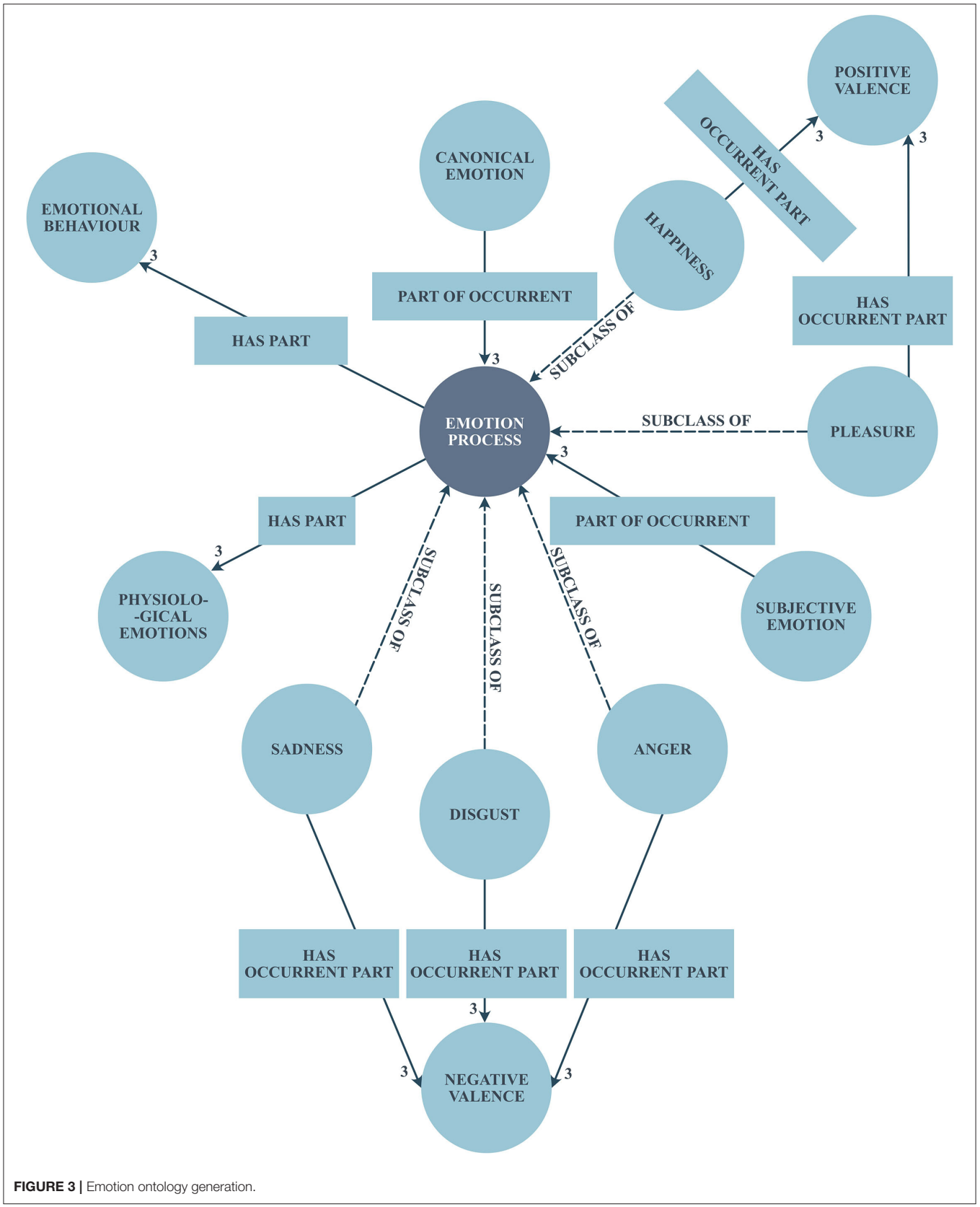


FIGURE 3 | Emotion ontology generation.

such as NoSlang Dictionary, Urban Dictionary, and Translit (44). Further, for every detected slang word, we generated the candidate lists based on the synsets produced from the WordNet Synset. We then used Dictionary.com, Thesaurus.com, and Oxford Dictionary for increasing the probability strength of the words. The intensity of the words resulting from these dictionaries would be considered and populated for our Emotion Ontology.

Emotion Detection in COVID-19 Datasets

Here, we provide a baseline approach that can be followed for automatically annotating the emotions extracted from the COVID-19 tweets[1]. To extract the emotions, we have applied the binary Support Vector Machine (SVM) classifier to effectively identify the eight emotions as proposed by Ekman (14). Each classifier has been trained independently for every emotion which resulted in seven independent binary classifiers and further the merger of these combinations of classifiers would be considered as a single multi-label classifier. Using this single multi-label classifier, the tweets can be annotated even if they possess more than one emotion and return positive if the multi-binary classifier is applied to the tweet. Likewise, it would return negative results if a binary classifier is applied. The result would be neutral for no emotions in the tweet. Each classifier has been following an independent classifier for a different group of features and in particular, the feature used by the binary classifier is a subset of the features deployed in the multi-label classifier (45).

To augment the feature extraction process accurately, we have omitted some of the similarity features which envisioned more on a topic than emotions and rejected the synsets followed by WordNet Affect features in this research. The decision has been taken not to use WordNet Affect for this research as normally the tweet does not possess sufficient synsets to cross-verify over the WordNet Affect. Moreover, WordNet Affect would probably look for possible candidate hypernyms instead of potential resources (i.e., thing, or object, in our case). Another dissimilarity observed on WordNet Affect is that it focused mainly on bigrams and trigrams for normalization instead of phrases. This would not be largely amicable for tweets due to their constraint in length. However, for the effective utilization of feature extraction for the proposed research, we used the topic modeling approach to discover the hidden terms and phrases between the tweets even though the tweet has no entities in common. Here, we have employed the LDA algorithm for topic modeling and it discovers the similarity that exists among the collections of tweets (i.e., the collection of tweets can be considered as a document by the LDA algorithm). For implementing this approach, we have enforced MALLET-based LDA implementation and treated every tweet as its document (46). This approach would project the probabilistic combination of words for every tweet and each topic is generalized into a set of probabilistic combinations of words.

Tweet Polarity Calculation Mechanism

The ultimate objective of this proposed approach is to calculate the polarity of every tweet that we have collected from COVID-19 datasets. In the conventional methods, the sentiment score will be calculated based on the word polarity and the overall strength will be estimated on the grounds of the polarity score obtained from the probability of the positive and negative words. But in this proposed approach, we have modeled a novel approach to get the polarity of words on every tweet by using the SPARQL query that crawls through the emotional ontology to fetch the correct sense for the search word. In our proposed model, we have excluded the pronouns, adjectives, and articles from the tweets and extracted the tokens such as nouns, verbs, objects, and determiners. As the Semantic Web gives a well-defined meaning to the sentences and converts the sentences into appropriate triples, such as SPO, we have followed the same pattern for our proposed approach for the Emotion Ontology building process and extracted the nouns (subjects), verbs (predicates), and objects (objects) that satisfy the Semantic Web triple pattern, respectively. The relation between the entities (either noun or object) can be expressed using the RDF/RDFS. The ontology population can be done either by the tool Protégé or using the Semantic Web language, OWL. For logics and inferences possessed in the ontology, we have used the first-order logics as well as description logics to express the relations given in the Emotion Ontology. Mostly, we have utilized the predicates for relationships between two or more entities and the predicate relationships can be expressed using the OWL predicate type method. For each tweet, we execute the SPARQL query to obtain the disambiguated result. The SPARQL query to fetch the entities embedded with emotions is given below: The Emotion Ontology helps to disambiguate

```
SELECT ?group ?emotion
WHERE
{
?a rdf:type x:EmotionAboutEntityItem;
x:hasEmotion ?emotion;
x:hasEntityItem ?entityItem.
?group x:contains ?entityItem.
}
ORDER BY ASC(?group)
```

the search terms accurately and to return the results precisely. Thereby, the polarity and sentiment score for each term or entity can be obtained and yield an accurate sentiment score for the tweets. Suppose, the search terms used for the execution in the SPARQL query return NIL results, then the context of the token would be considered and an appropriate parameterized SPARQL query will be executed against the Emotion Ontology to fetch the right mix of words for consideration. We then estimated the sentiment score by considering the largest domain value from the list of positive tokens and the smallest values from the other set of negative tokens. Eventually, to determine the polarity of every tweet, the tweet sentiment score is calculated based on the strength score of the tweet, and the strength score of every entity in the tweet is calculated based on the following calculations: (a) positive, if sentiscore is greater than 1, (b) negative if sentiscore is less than or equal to -1 and (c) otherwise neutral.

TABLE 3 | Tweet segmentation and classification for ontology matching.

Segmentation	Precision	Recall	F-Score
Tweet Seg + Ontology	0.892	0.801	0.841
Tweet Seg + Inferences	0.761	0.684	0.692
POS + First Order Logics	0.561	0.504	0.539
Tweet Seg + POS	0.710	0.693	0.703

TABLE 4 | Total number of occurrences of the emotions tested for the training dataset.

Joy	Anger	Disgust	Fear	Surprise	Love	Sadness	Confusion
187	562	491	688	490	294	579	463

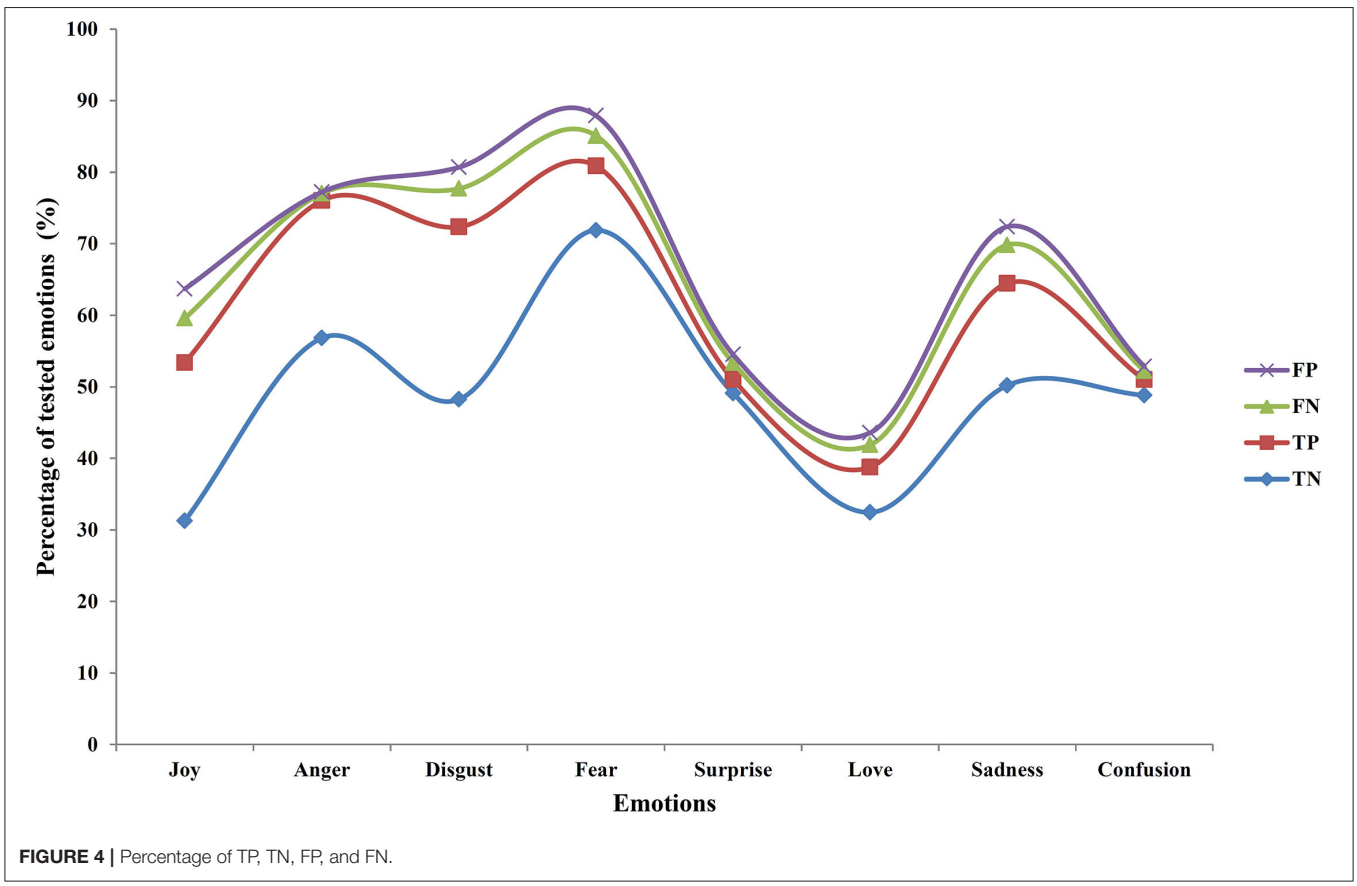


FIGURE 4 | Percentage of TP, TN, FP, and FN.

TABLE 5 | Calculation of the percentage of TN, FP, FN, and TP from the confusion matrix.

	Joy	Anger	Disgust	Fear	Surprise	Love	Sadness	Confusion
TN	31.27	56.81	48.24	71.88	49.11	32.45	50.19	48.82
TP	22.10	19.21	24.11	8.99	1.90	6.33	14.28	2.19
FN	6.21	1.01	5.35	4.22	2.34	3.11	5.36	1.21
FP	4.10	0.19	2.98	2.81	1.16	1.69	2.55	0.61

Further, to assess the absolute performance of our proposed Emotion Ontology, we have considered two implicit functions, such as sklearn.metrics.recall and sklearn.metrics.f1 to calculate the potential values, such as recall and F1 score, respectively. These two metrics were used to handle the multi-label classification (positive, negative, and neutral). Our sheer Emotion

TABLE 6 | Emotion classifier for each emotion extracted from COVID-19 datasets.

Emotion	#Tweets	Precision	Recall	F-score
Joy	273	0.781	0.689	0.735
Anger	849	0.674	0.621	0.647
Disgust	783	0.738	0.685	0.711
Fear	959	0.811	0.783	0.797
Surprise	630	0.893	0.812	0.852
Love	384	0.874	0.823	0.848
Sadness	791	0.891	0.841	0.866
Confusion	768	0.783	0.738	0.760
Macro-Average	817	0.805	0.749	0.777

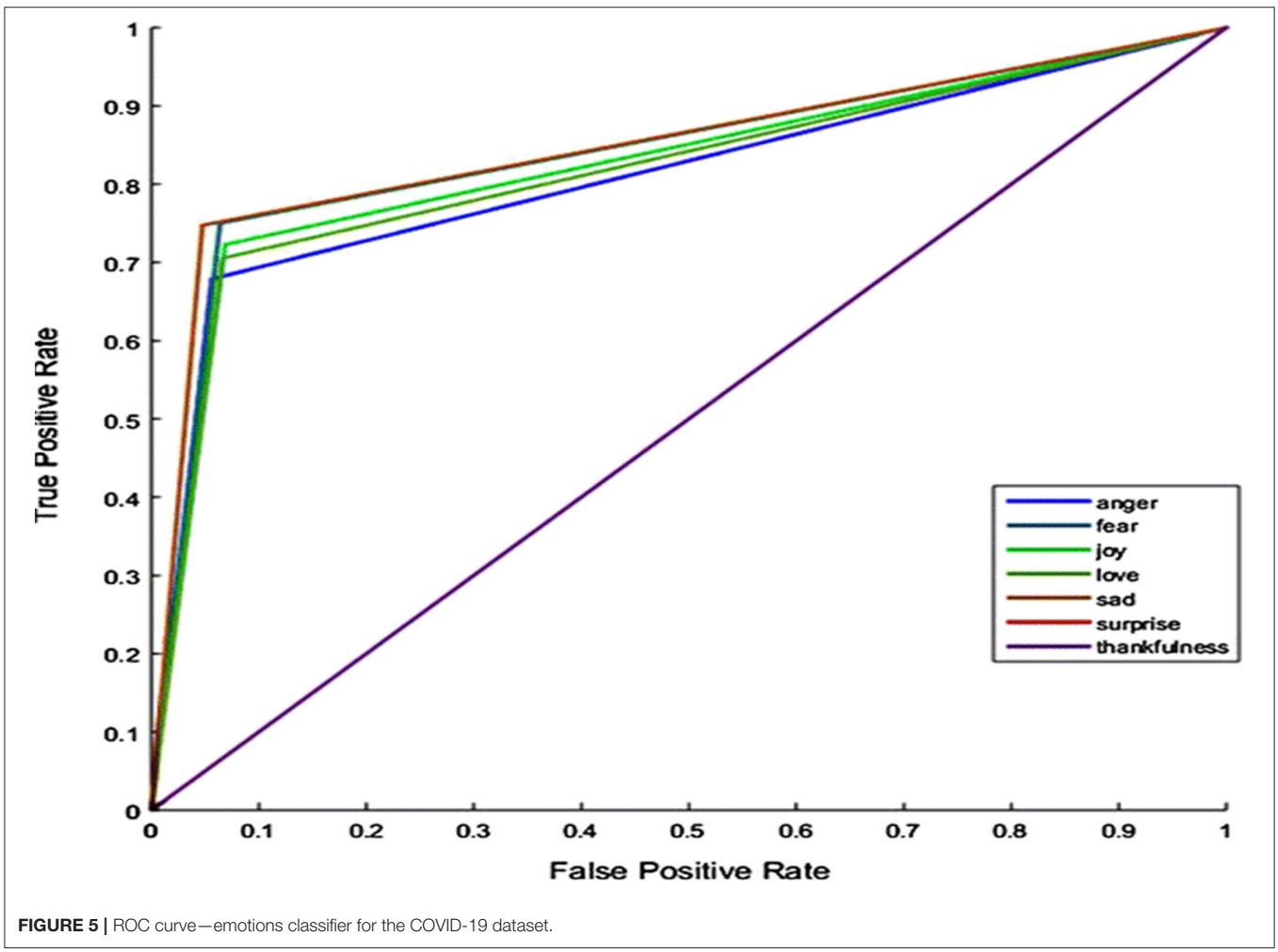


FIGURE 5 | ROC curve—emotions classifier for the COVID-19 dataset.

Ontology population using Semantic Web technologies and performance measures selected for sentiment score has enabled the proposed method to avoid manual annotation using any machine learning techniques, and the problem of manual annotation has been resolved. Further, many recent studies have sought the help of machine learning algorithms to manually annotate the entities given in the text or any tweets. The proposed Emotion Ontology can deal with dynamic domain vocabularies that would possess different representations in the text.

Experimental Evaluation of Emotion Ontology Matching

Table 3 shows the performance of our Emotion Ontology and categorizes the accuracy rate based on the different segmentation

TABLE 7 | Comparative analysis with existing models and their limitations.

Reference	Year	Approach	Data sets	One sentence summary	Limitations	F1-score
Our Proposed Model	–	Ontological Framework	COVID-19 Pandemic related Tweets	Right utilization of Ontological Framework coupled with ML algorithms such as SVM and LDA	Can work only with English Text.	0.831
Matla and Badugu (47)	2020	Machine Learning	Tweets	Deployed the Naive Bayes algorithm for classifying the Twitter messages into four emotional categories.	Absences of contextual words in the text.	0.725
Huang et al. (48)	2019	Visual and semantic attention mechanism	Tweets	Deep Multimodal Attentive Fusion for multimodal sentiment analysis of 10 million Tweets.	The fine-granularity relation between image and text pairs is not yet explored.	0.769
Ragheb et al. (49)	2019	Machine Learning	CLEF eRisk-2019 (signs of anorexia) T1 dataset	Categorized the emotions based on Attention based model.	Cannot detect the happy emotions from text.	0.710
Hasan et al. (50)	2019	Hybrid	Tweets	Deployed a supervised learning system for automatically classifying emotion in text stream messages using ANEW lexicon.	Works well only for text messages.	0.778
Almanie et al. (51)	2018	Rule Based	Tweets	Able to extract the prominent 5 emotions in a real time situation.	Low semantic orientation of textual context.	0.719
Badugu and Suhasini. (6)	2017	Rule Based	Tweets	Detected only 4 primary emotions using the Rule Based approach.	This system focused only on English sentences.	0.720

followed for tweet analysis. Recall of the proposed ontology pinpoints the various expression and candidate terms employed over the tweets and reduced the false-positive cases encountered. The tweet segmentation has been carried out for tweets, part-of-speech (POS) tagging, and the likelihood of context rule applied for the analysis.

From the COVID-19 datasets retrieved from Twitter, training was strenuously performed over all the sample tweets, and validation was carried out with the baseline gold standard datasets. The total number of occurrences for each emotion listed was calculated by the number of positive values on the respective emotions and identified from class imbalances over the emotions, such as joy and love. **Table 4** outlines the occurrences of each emotion tested for the COVID-19 dataset.

To measure the accuracy of the proposed system, it has always relied on the precise calculation of the confusion matrix and mostly it is to obtain a low score over false positives (FP). Our proposed work is able to extract and identify almost 81% of true positives (TP) but it has encountered some serious challenges in detecting the false negatives (FN). This result has been vividly depicted in **Table 5** that projects the amount of FP generated against FN. **Figure 4** portrays the percentage of TP, true negatives (TN), FP, and FN.

Eventually, each core emotion that was extracted from the tweets have been analyzed in terms of precision, recall, and f-score, and the details are provided in **Table 6**.

Our proposed model considers every emotion with its associated context returned from the Emotion Ontology, and the sentiment score has been calculated based on the sum, average, and maximum count of emotion score returned by the Emotion Ontology. It has been witnessed in the analysis that our proposed system has extracted the emotional words and their associated context through our proposed Emotional Ontology. **Figure 5** illustrates the receiver operating characteristic curve (ROC) of the emotion classifier for the COVID-19 dataset.

Our proposed model has been compared with some of the benchmarked baseline approaches to highlight the major differences established in terms of F1-score. Some of the limitations that have been witnessed in the existing approaches have been delineated in **Table 7**. Major disadvantages of the existing approaches were mostly due to the lack of contextual information, low semantic orientation between two words, and availability of limited categories to extract the emotions.

CONCLUSION

The proposed Emotion Ontology model has been robust and efficient in extracting the full range of human emotions pertaining to COVID-19-related concepts. The emotions related to COVID-19 have been obtained and several emotional expressions, such as intensifiers, negator, lexical features, interjections, and conjunctions were considered for populating the Emotion Ontology. Semantic Web technologies have been utilized for creating the triples for every tweet, and the ontology construction has been manually done with the use of OWL functions. Compared to other lexicon-based emotion methods and machine learning approaches, the performance of

the proposed system is satisfactory. As the tweets have been considerably informal in nature and have sparse content, in particular, the other lexical-based model and machine learning approaches failed to yield the desired sentiment score and ended up with some satisfactory precision and recall score. Our proposed study is able to extract and detect almost 81% of TP and is able to detect a considerable amount of FN. Since this study mainly focused on dealing with the English language, there is a certain limitation for extracting few emotions that exist in other languages, such as French, Spanish, etc. For future work, we will consider extracting the profound availability of emotional occupancy of resources in other languages and level their accuracy on par with standard metrics defined in the baseline approach.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SN and C-YC did the conceptualization and supervised the research. C-YC carried out the funding acquisition. SN investigated the data, performed the methodology, and implemented the software code. KS and C-YC carried out the project administration and validated the results. SN and KS wrote the manuscript. SN, KS, and SM reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially funded by Intelligent Recognition Industry Service Research Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and Ministry of Science and Technology in Taiwan (Grant No. MOST 109-2221 -E-224-048-MY2).

REFERENCES

1. Al-Aswadi FN, Chan HY, Gan KH. Automatic ontology construction from text: a review from shallow to deep learning trend. *Artific Intell Rev.* (2020) 53:3901–28. doi: 10.1007/s10462-019-09782-9
2. Chakriswaran P, Vincent DR, Srinivasan K, Sharma V, Chang, C.-Y.; Reina, D.G. Emotion ai-driven sentiment analysis: a survey, future research directions, open issues. *Appl. Sci.* 9:5462. doi: 10.3390/app9245462
3. Lin C, Ibeke E, Wyner A, Guerin F. Sentiment–topic modeling in text mining. *Wiley Interdisciplin Rev Data Mining Knowl Discov.* (2015) 5:246–54. doi: 10.1002/widm.1161
4. Srinivasan K., Garg L., Datta D., Alaboudi A. A., Jhanjhi N. Z., et al. Performance comparison of deep cnn models for detecting driver's distraction. *Comput Mater Continua.* (2021) 68:4109–24. doi: 10.32604/cmc.2021.016736
5. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-technological factors affecting user's adoption of ehealth functionalities: a case study of China and ukraine ehealth systems. *IEEE Access.* (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
6. Badugu S, Suhasini M. Emotion detection on twitter data using knowledge base approach. *Int J Comput Appl.* (2017) 162:366. doi: 10.5120/ijca2017913366
7. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Front. Public Health* 8:357. doi: 10.3389/fpubh.2020.00357
8. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimed Syst.* (2021) 28:1–15. doi: 10.1007/s00530-021-00774-w
9. Jayalakshmi M, Garg L, Maharajan J, Jayakumar KK, Srinivasan K. et al. Fuzzy logic-based health monitoring system for covid'19 patients. *Comput Mater Continua.* (2021) 67:2431–47. doi: 10.32604/cmc.2021.015352
10. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Mathematic Biosci Eng.* (2021) 14:18. doi: 10.3934/mbe.2021418
11. Krishnamurthy S, Srinivasan K, Qaisar SM, Vincent PM, Chang CY. Evaluating deep neural network architectures with transfer learning for pneumonitis diagnosis. *Comput Math Meth Med.* (2021) 12:8036304. doi: 10.1155/2021/8036304
12. Ali F, Kwak KS, Kim YG. Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification. *Appl Soft Comput.* (2019) 47:235–50. doi: 10.1016/j.asoc.2016.06.003
13. Lang PJ. A bio-informational theory of emotional imagery. *Psychophysiology.* (1979) 16:495–512. doi: 10.1111/j.1469-8986.1979.tb01511.x
14. Ekman P. Expression and the nature of emotion. *Approaches to emotion.* (1984) 3:344.
15. Scherer KR, Dieckmann A, Unfried M, Ellgring H, Mortillaro M. Investigating appraisal-driven facial expression and inference in emotion communication. *Emotion.* (2021) 21:73–95. doi: 10.1037/emo0000693
16. Lacy LW. *OWL: Representing Information using the Web Ontology Language.* https://www.google.com/search?client=firefox-b-d&q=Bloomington+Indiana&stick=H4sIAAAAAAAAAOPgE-LUz9U3SDEtzDBS4gAxi_NKKrW0spOt9POL0hPzMqsSSzLz81A4VhmpiSmFpYlFJalFxYtYRZxy8_vNzM_PSS_LzdBQ881IyE_MSd7AyAgDbctB_XQAAAA&sa=X&ved=2ahUKEwjQseaujp_0AhVahHIEHc9CAKoQmxMoAXoECEIQAw Bloomington, IN: Trafford Publishing (2005).
17. Baldauf M, Dustdar S, Rosenberg F. A survey on context-aware systems. *Int J Ad Hoc Ubiquitous Comput.* (2007) 2:263–277. doi: 10.1504/IJAHUC.2007.014070
18. Thompson JJ, Leung BH, Blair MR, Taboada M. Sentiment analysis of player chat messaging in the video game StarCraft 2: extending a lexicon-based model. *Knowl Based Syst.* (2017) 137:149–62. doi: 10.1016/j.knosys.2017.09.022
19. Ren G, Hong T. Investigating online destination images using a topic-based sentiment analysis approach. *Sustainability.* (2017) 9:1765. doi: 10.3390/su9101765
20. Oliveira N, Cortez P, Areal N. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Syst Appl.* (2017) 73:125–44. doi: 10.1016/j.eswa.2016.12.036
21. Ren R, Wu DD, Liu T. Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Syst J.* (2018) 13:760–770. doi: 10.1109/JSYST.2018.2794462
22. Baldoni M, Baroglio C, Patti V, Rena P. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale.* (2017) 6:41–54. doi: 10.3233/IA-2012-0028
23. Dey A, Jenamani M, Thakkar JJ. Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Syst Appl.* (2018) 103:92–105. doi: 10.1016/j.eswa.2018.03.004

24. Gerber D, Hellmann S, Bühmann L, Soru T, Usbeck R, Ngomo ACN. (2013) "Real-time RDF extraction from unstructured data streams," In: *International Semantic Web Conference*. Berlin, Heidelberg: Springer, p. 135–150.
25. Madani Y, Erritali M, Bengourram J. Sentiment analysis using semantic similarity and Hadoop MapReduce. *Knowl Inform Syst.* (2019) 59:413–36. doi: 10.1007/s10115-018-1212-z
26. Mozafari F, Tahayori H. "Emotion detection by using similarity techniques," In: *2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)* Piscataway, NJ: IEEE (2019). p. 1–5.
27. Dridi A, Recupero DR. Leveraging semantics for sentiment polarity detection in social media. *Int J Mach Learn Cybern.* (2019) 10:2045–55. doi: 10.1007/s13042-017-0727-z
28. Dragoni M, Poria S, Cambria E. OntoSenticNet: A commonsense ontology for sentiment analysis. *IEEE Intell Syst.* (2018) 33:77–85. doi: 10.1109/MIS.2018.033001419
29. Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Syst Appl.* (2017) 72:221–30. doi: 10.1016/j.eswa.2016.10.065
30. Li Y, Pan Q, Yang T, Wang S, Tang J, Cambria E. Learning word representations for sentiment analysis. *Cogn Computat.* (2017) 9:843–851. doi: 10.1007/s12559-017-9492-2
31. Zhu G, Iglesias CA. Computing semantic similarity of concepts in knowledge graphs. *IEEE Trans Knowl Data Eng.* (2016) 29:72–85. doi: 10.1109/TKDE.2016.2610428
32. Bruni E, Tran NK, Baroni M. Multimodal distributional semantics. *J Artific Intell Res.* (2014) 49:1–47. doi: 10.1613/jair.4135
33. Hill F, Korhonen A. (2014) "Learning abstract concept embeddings from multi-modal data: Since you probably can't see what I mean," In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP), p. 255–65.
34. Deng S, Sinha AP, Zhao H. Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Syst.* (2017) 94:65–76. doi: 10.1016/j.dss.2016.11.001
35. Kumar NS, Muruganantham D. Disambiguating the twitter stream entities and enhancing the search operation using dbpedia ontology: named entity disambiguation for twitter streams. *Int J Inform Technol Web Eng.* (2016) 11:51–63. doi: 10.4018/IJITWE.2016040104
36. Strapparava C, Valitutti A. Wordnet affect: an affective extension of wordnet. *In Lrec.* (2004) 4:1083–6. http://www.lrec-conf.org/proceedings/lrec2004/pdf/369.pdf
37. Mostafa MM. More than words: Social networks' text mining for consumer brand sentiments. *Expert Syst Appl.* (2013) 40:4241–51. doi: 10.1016/j.eswa.2013.01.019
38. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining. *LREc.* (2010) 10:1320–26. http://www.lrec-conf.org/proceedings/lrec2010/pdf/385_Paper.pdf
39. Shadbolt N, Berners-Lee T, Hall W. The semantic web revisited. *IEEE Intell Syst.* (2006) 21:96–101. doi: 10.1109/MIS.2006.62
40. Li S, Liu L, Xiong Z. Ontology-based sentiment analysis of network public opinions. *Int J Digit Cont Technol Appl.* (2012) 6:371. doi: 10.4156/jdcta.vol6.issue23.42
41. Ghiassi M, Skinner J, Zimbra D. Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst Appl.* (2013) 40:6266–82. doi: 10.1016/j.eswa.2013.05.057
42. Lim MY, Aylett R. An emergent emotion model for an affective mobile guide with attitude. *Appl Artific Intell.* (2009) 23:835–54. doi: 10.1080/08839510903246518
43. Francisco V, Hervás R, Peinado F, Gervás P. EmoTales: creating a corpus of folk tales with emotional annotations. *Lang Resourc Eval.* (2012) 46:341–81. doi: 10.1007/s10579-011-9140-5
44. Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *J Comput Sci.* (2011) 2:1–8. doi: 10.1016/j.jocs.2010.12.007
45. van den Broek-Altenburg EM, Atherly AJ. Using social media to identify consumers' sentiments towards attributes of health insurance during enrollment season. *Appl Sci.* (2019) 9:2035. doi: 10.3390/app9102035
46. Zhang Y, Song D, Zhang P, Li X, Wang P. A quantum-inspired sentiment representation model for twitter sentiment analysis. *Appl Intell.* (2019) 49:3093–3108. doi: 10.1007/s10489-019-01441-4
47. Suhasini M, Srinivasu B. "Emotion detection framework for twitter data using supervised classifiers," In: *Data Engineering and Communication Technology*. Singapore: Springer. (2020) p. 565–76.
48. Huang F, Zhang X, Zhao Z, Xu J, Li Z. Image–text sentiment analysis via deep multimodal attentive fusion. *Knowl Based Syst.* (2019) 167:26–37. doi: 10.1016/j.knosys.2019.01.019
49. Ragheb W, Azé J, Bringay S, Servajean M. (2019) Attentive multi-stage learning for early risk detection of signs of anorexia and self-harm on social media. In: *CLEF (Working Notes)*.
50. Hasan M, Rundensteiner E, Agu E. Automatic emotion detection in text streams by analyzing Twitter data. *Int J Data Sci Anal.* (2019) 7:35–51. doi: 10.1007/s41060-018-0096-z
51. Almanie T, Aldayel A, Alkanhal G, Alesmail L, Almutlaq M, Althunayan R. (2018) "Saudi Mood: a real-time informative tool for visualizing emotions in Saudi Arabia Using Twitter," In: *2018 21st Saudi Computer Society National Computer Conference (NCC)*. Piscataway, NJ : IEEE. p. 1–6.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Narayanasamy, Srinivasan, Mian Qaisar and Chang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fpubh.2021.812023

Machine Learning Applications for the Prediction of Bone Cement Leakage in Percutaneous Vertebroplasty

Wenle Li[1,2†], Jiaming Wang[3†], Wencai Liu[4†], Chan Xu[2,5], Wanying Li[2], Kai Zhang[1,2], Shibin Su[6], Rong Li[7], Zhaohui Hu[8], Qiang Liu[1*], Ruogu Lu[9*] and Chengliang Yin[10*]*

[1] Department of Orthopedics, Xianyang Central Hospital, Xianyang, China, [2] Clinical Medical Research Center, Xianyang Central Hospital, Xianyang, China, [3] Department of Orthopedics, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [4] Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, China, [5] Department of Dermatology, Xianyang Central Hospital, Xianyang, China, [6] Xiamen Bank, Xiamen, China, [7] The First Affiliated Hospital, Shaanxi University of Traditional Chinese Medicine, Xianyang, China, [8] Department of Spine Surgery, Liuzhou People's Hospital, Liuzhou, China, [9] Department of Electronic and Computer Science, University of Southampton, Southampton, United Kingdom, [10] Faculty of Medicine, Macau University of Science and Technology, Macau, China

Background: Bone cement leakage is a common complication of percutaneous vertebroplasty and it could be life-threatening to some extent. The aim of this study was to develop a machine learning model for predicting the risk of cement leakage in patients with osteoporotic vertebral compression fractures undergoing percutaneous vertebroplasty. Furthermore, we developed an online calculator for clinical application.

Methods: This was a retrospective study including 385 patients, who had osteoporotic vertebral compression fracture disease and underwent surgery at the Department of Spine Surgery, Liuzhou People's Hospital from June 2016 to June 2018. Combing the patient's clinical characteristics variables, we applied six machine learning (ML) algorithms to develop the predictive models, including logistic regression (LR), Gradient boosting machine (GBM), Extreme gradient boosting (XGB), Random Forest (RF), Decision Tree (DT) and Multilayer perceptron (MLP), which could predict the risk of bone cement leakage. We tested the results with ten-fold cross-validation, which calculated the Area Under Curve (AUC) of the six models and selected the model with the highest AUC as the excellent performing model to build the web calculator.

Results: The results showed that Injection volume of bone cement, Surgery time and Multiple vertebral fracture were all independent predictors of bone cement leakage by using multivariate logistic regression analysis in the 385 observation subjects. Furthermore, Heatmap revealed the relative proportions of the 15 clinical variables. In bone cement leakage prediction, the AUC of the six ML algorithms ranged from 0.633 to 0.898, while the RF model had an AUC of 0.898 and was used as the best performing ML Web calculator (https://share.streamlit.io/liuwencai0/pvp_leakage/main/pvp_leakage) was developed to estimate the risk of bone cement leakage that each patient undergoing vertebroplasty.

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Zhaoyang Han,
University of Aizu, Japan
Huajing Teng,
Beijing Institutes of Life Science (CAS), China

***Correspondence:**
Chengliang Yin
chengliangyin@163.com
Ruogu Lu
15622769758@163.com
Qiang Liu
m13992079668@163.com
Zhaohui Hu
329870376@qq.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 November 2021*
Accepted: *17 November 2021*
Published: *10 December 2021*

Citation:
Li W, Wang J, Liu W, Xu C, Li W, Zhang K, Su S, Li R, Hu Z, Liu Q, Lu R and Yin C (2021) Machine Learning Applications for the Prediction of Bone Cement Leakage in Percutaneous Vertebroplasty. Front. Public Health 9:812023. doi: 10.3389/fpubh.2021.812023

Conclusion: It achieved a good prediction for the occurrence of bone cement leakage with our ML model. The Web calculator concluded based on RF model can help orthopedist to make more individual and rational clinical strategies.

Keywords: percutaneous vertebroplasty, bone cement leakage, machine learning algorithms, prediction model, web calculator

INTRODUCTION

Percutaneous vertebroplasty (PVP) was a minimally invasive technique that uses liquid polymethylmethacrylate bone cement injected into the fractured vertebrae to reduce pain, strengthen the bone and prevent further compression of the vertebral body. PVP has become the mainstay treatment for osteoporotic and malignant vertebral fractures. Leakage of bone cement into the intervertebral disc increases the risk of adjacent vertebral fractures paraplegia due to extravasation of bone cement into the epidural or spinal cord compression associated with the intervertebral foramen (1). Besides, Leakage of bone cement into the paravertebral vein can lead to pulmonary embolism cardiac perforation, cerebral embolism and even death (2). Meanwhile, there are few reports of screening for independent risk factors and risk prediction models for bone cement leakage in patients undergoing orthopaedic percutaneous vertebroplasty.

Algorithmic decision support tools, computer-assisted navigation and surgical robots have served in the clinic (3, 4). Machine learning (ML) has been widely used in healthcare data analysis and some in spinal surgery (5, 6). By the comparison among ML algorithms, a better predictive model can be constructed to predict cement leakage after percutaneous vertebroplasty.

Some studies elaborated that age, bone mineral density (BMD), body mass index (BMI), volume of bone cement injections, and other underlying chronic diseases were risk factors for cement leakage (7). In addition, we added injury-to-operation time, hospitalization-to-operation time, and length of surgery.

Our study aimed to obtain the best ML algorithm, and constructed a big data risk prediction web calculator for postoperative cement leakage in patients undergoing percutaneous vertebroplasty. It ultimately provided a basis for future treatment and prevention strategies related to postoperative complications.

METHODS

Study Population

All data for this study were available from the Department of Spine Surgery, Liuzhou People's Hospital. This study recorded the medical variables of patients hospitalized with or without cement leakage after undergoing percutaneous vertebroplasty, from June 2016 to June 2018. Retrospectively analyzed the basic patient's information, as well as baseline clinical characteristics, surgical details.

Data Collection

To develop the target machine learning model, several variables were obtained, including patient age, height, weight, body mass index, bone mineral density, volume of bone cement injection, length of hospital stay to surgery, length of surgery, history of previous disease, and medication, as well as whether re-fracture occurred after surgery. Various machine learning algorithms such as logistic regression, random forest and other currently popular machine learning models were utilized in this study. Inclusion criteria for inpatient data were as follows: (1) osteoporotic vertebral fracture with moderate or severe low back pain which limited movement as the main complaint; (2) no significant improvement after active conservative treatment; (3) one or more vertebral fractures confirmed by imaging; (4) bone mineral density <2.5 by dual energy X-ray bone densitometry. In addition, exclusion criteria were: (1) fractures with symptoms of spinal nerve compression; (2) pathological vertebral fractures, such as spinal tumors, vertebral hemangiomas, spinal tuberculosis, etc.; and (3) patients with poor physical fitness and intolerance to surgical treatment. This study was approved by the Institutional Review Board of the Liuzhou People's Hospital before data collection and analysis, the IRB approval number is 2020 (KY-E-22-01).

Surgical Strategy

Patients undergo individualized PVP in the prone position with local anesthesia. All PVPs were performed with a unilateral lateral approach through the vertebral arch. The puncture was performed through the vertebral arch under fluoroscopic guidance. The puncture point is located at the intersection of a cephalolateral with a horizontal line 5–15 mm from the arch projection and a lateral vertical line $\sim$ 30 mm from the posterior midline. Under fluoroscopy, a 3.5 mm needle was moved along the root of the arch, the head adjustment and abduction angle of the needle bundle was allowed until the junction of the anterior third of the vertebral body was reached. The spread and distribution of the bone cement was monitored under fluoroscopy while the patient was tested for bilateral lower limb movement. After that, needle was withdrawn, covered with a sterile dressing and secured with pressure. Pre- and post-operative imaging is routinely performed in screened cases to determine the degree of compression of the fractured vertebral body, the height of the vertebral body, the integrity of the posterior wall of the vertebral body, the amount of cement

Abbreviations: ML, machine learning; LR, logistic regression; GBM, Gradient boosting machine; XGB, Extreme gradient boosting; RF, Random Forest; DT, Decision Tree; MLP, Multilayer perceptron; AUC, Area Under Curve; PVP, Percutaneous vertebroplasty; BMD, bone mineral density; BMI, body mass index; OR, odds ratio; CI, confidence intervals.

TABLE 1 | Baseline table of patients with and without cement leakage.

Variables	level	Overall (*N* = 385)	No (*N* = 304)	Yes (*N* = 81)	*p*-value
Age (median [IQR])	NA	75.400 [68.300, 80.700]	75.650 [68.300, 80.750]	74.900 [68.200, 80.600]	**0.987**
High (median [IQR])	NA	155.000 [149.000, 161.000]	155.000 [148.000, 160.000]	155.000 [149.000, 162.000]	**0.2082**
Weigh (median [IQR])	NA	47.000 [40.000, 58.000]	47.000 [41.000, 58.000]	47.000 [40.000, 60.000]	**0.948**
BMI (median [IQR])	NA	19.899 [17.116, 23.873]	19.905 [17.360, 23.905]	19.819 [16.437, 23.613]	**0.4335**
BMD (median [IQR])	NA	4.400 [3.900, 5.000]	4.400 [3.900, 5.000]	4.600 [4.100, 4.900]	**0.1833**
Injection volume of bone cement (median [IQR])	NA	4.000 [3.500, 5.000]	4.000 [3.500, 5.000]	4.000 [4.000, 5.000]	**0.0558**
Hospital stay to surgery (median [IQR])	NA	5.000 [4.000, 6.000]	5.000 [3.000, 6.000]	5.000 [4.000, 7.000]	**0.1086**
Injury to surgery (median [IQR])	NA	14.000 [8.000, 30.000]	14.000 [8.000, 30.000]	16.000 [10.000, 33.000]	**0.2193**
Antiosteoporosis (%)	No	245 (63.64)	192 (63.16)	53 (65.43)	**0.804**
	Yes	140 (36.36)	112 (36.84)	28 (34.57)	
Multiple (%)	No	205 (53.25)	177 (58.22)	28 (34.57)	**0.0002**
	Yes	180 (46.75)	127 (41.78)	53 (65.43)	
Steroid (%)	No	320 (83.12)	248 (81.58)	72 (88.89)	**0.1634**
	Yes	65 (16.88)	56 (18.42)	9 (11.11)	
Re-fracture (%)	No	327 (84.94)	260 (85.53)	67 (82.72)	**0.6502**
	Yes	58 (15.06)	44 (14.47)	14 (17.28)	

injected intraoperatively, whether the cement leaks, and whether the vertebral body is re-fractured postoperatively.

Statistical Analyses

Exploring the independent risk factors for PVP, multivariate logistic regression analysis was applied to calculate the odds ratio (OR) with 95% confidence intervals (CI) using a backward stepwise selection method. The value of OR >1 indicates that the variable is a risk factor. $P < 0.05$ was regarded as statistically significant. The predictive ability of ML algorithms has outperformed traditional regression methods in large databases (8). We used six different ML algorithms to analyze our data: logistic regression (LR), Gradient boosting machine (GBM), Extreme gradient boosting (XGB) Random Forest (RF), Decision Tree (DT) and Multilayer perceptron (MLP). After training, the one with the highest average AUC was chosen as the best algorithm. Furthermore, the ML-based model was tuned to avoid overfitting and the accuracy of the algorithm was tested using ten-fold cross-validation method.

In our study, the components and proportions of multiple relevant factors related to PVP were analyzed and visualized by heat map for the collected patient data. Furthermore, the heat map can also be used to identify data of outrageous quality during the actual analysis and assist in quality control.

We first used a multiple regression model as a benchmark for data evaluation and ML (which can handle a large number of input variables, assess the importance of variables when deciding on categories, and detect variable interactions). In the performance comparison of ML algorithms, the more AUC close to one, the better performance these classification model obtained. Based on the best execution model, a web calculator was created in this paper which implemented the prediction of illness for inputting patient data. The calculator allowed the surgeon to assess the risk of possible postoperative cement leakage in patients undergoing percutaneous vertebroplasty.

RESULTS

Demographics Features

The clinical history of 385 patients who underwent PVP was summarized in this study, the number of patients who developed post-operative cement leakage was 81 and the number of patients who did not develop cement leakage was 304 (**Table 1**). Among these observed subjects, the median age was 75.4 years (IQR:68.3, 80.7), the median height of these patients was 155 cm (IQR:149,161), the median weight of the patients was 47kg (IQR:40,58). The median BMD was 4.4 (IQR:3.9,5.0). The *P* values of these indicators were 0.987, 0.2082, 0.948, and 0.1833, respectively. None of them were statistically different. It worth noting that the 205 patients (53.25%) who did not have multiple vertebral fractures ($P < 0.05$). Besides, no cement leakage occurred in 177 patients (58.22%) and cement leakage occurred in 28 patients (34.57%).

In the correlation heat map (**Figure 1**), 15 highly correlated features were chosen as predictors. Weight was absolutely correlated with BMI and to some extent with height. As BMI got a better indicator of obesity, we only considered BMI rather than height or weight.

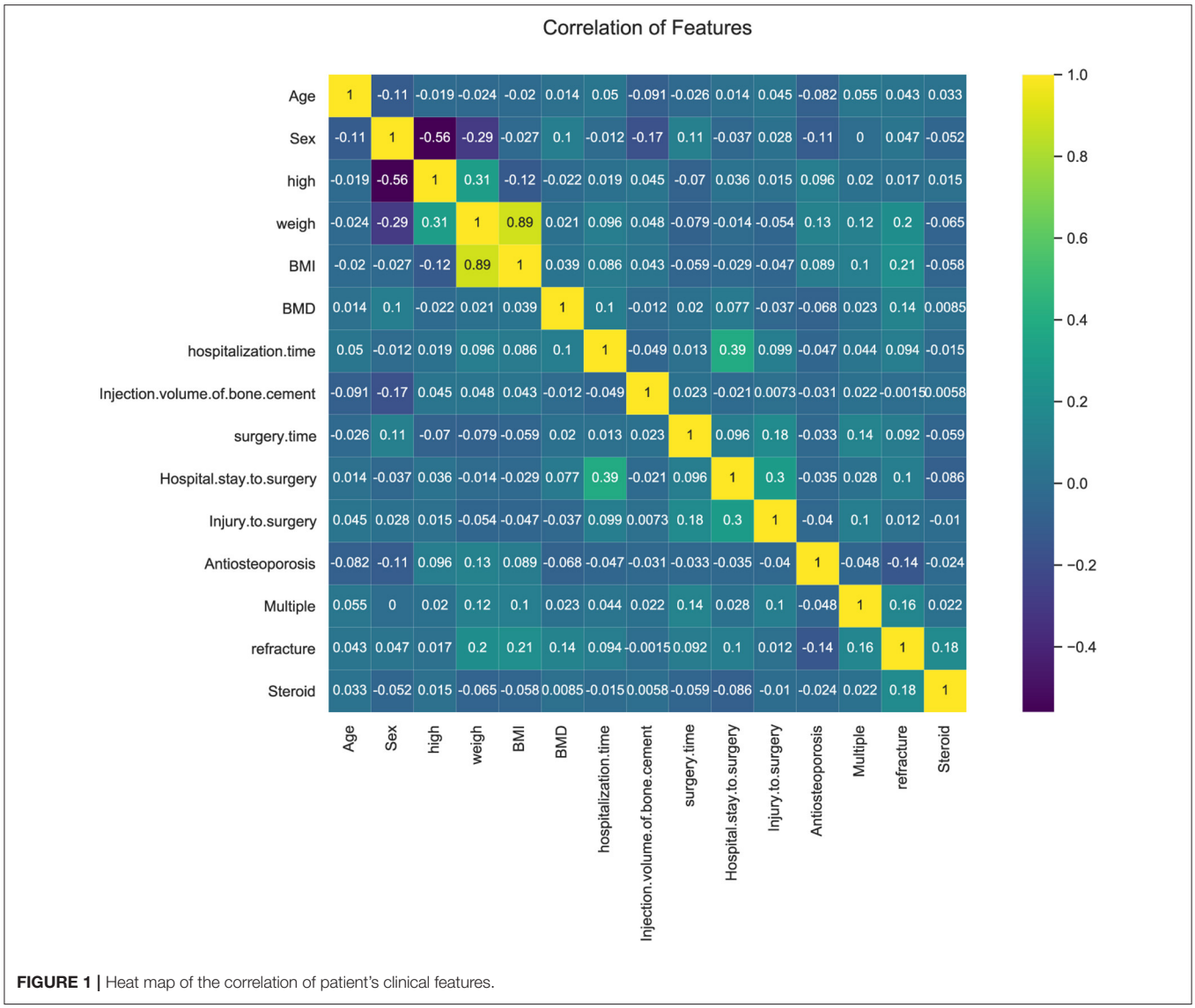


FIGURE 1 | Heat map of the correlation of patient's clinical features.

Univariate and Multivariate Logistic Regression Analyses of Cement Leakage

In terms of univariate analyzing (**Table 2**), age, height, weight, body mass index, bone mineral density, volume of bone cement injected, length of hospital stay, and intraoperative injury didn't have significant difference in occurrence of bone cement leakage in the overall population ($p > 0.05$).

In the multivariate logistic regression analysis, all parameters from the univariate analysis above were included. OR value showed the relative risk of bone cement leakage. The results showed that age (OR 1.002, 95% confidence interval 0.976–1.029), male (OR 0.799, 95% confidence interval 0.421–1.514), body mass index (OR 0.992, 95% confidence interval 0.946–1.040), bone mineral density (OR 1.272, 95% confidence interval 0.878–1.844), and length of hospital stay (OR 1.021, 95% confidence interval 0.973–1.072). Injection volume of bone cement, Surgery time, and Multiple vertebral fracture were all independent predictors for bone cement leakage. Besides, Multiple vertebral fractures (OR 2.638, 95% confidence interval 1.581–4.399), indicating a statistical difference ($P < 0.05$).

Performance of Machine Learning Algorithms

To compare the predictive performance of these six different ML algorithm models, 10-fold cross validation was applied in this study (**Figure 2**). As shown in the figure, the random forest model performed best in predicting bone cement leakage with an average AUC of 0.898. The remaining models were sorted in descending order depending on prediction performance. Therefore, we finally regarded the RF model as the preferred prediction model.

TABLE 2 | Univariate and multivariate logistic regression of bone cement leakage.

Variables	Univariate OR (95% CI)	*P*-value	Multivariate OR (95% CI)	*p*-value
Age (years)	1.002 (0.976–1.029)	**0.868**	/	**/**
Sex				
Female	Ref	**Ref**	/	**/**
Male	0.799 (0.421–1.514)	**0.492**	/	**/**
BMI	0.992 (0.946–1.040)	**0.757**	/	**/**
BMD	1.272 (0.878–1.844)	**0.202**	/	**/**
Hospitalization time (days)	1.021 (0.973–1.072)	**0.393**	/	**/**
Injection.volume.of.bone.cement (ml)	1.280 (1.0042–1.633)	**0.046**	1.283 (1.004-1.639)	**0.046**
Refracture				
No	Ref	**Ref**	/	**/**
Yes	1.234 (0.639–2.385)	**0.530**	/	**/**
Surgery time (min)	1.016 (1.004–1.028)	**0.005**	1.014 (1.002-1.027)	**0.017**
Hospital.stay.to.surgery (days)	1.058 (0.983–1.140)	**0.130**	/	**/**
Injury to surgery (days)	0.998 (0.992–1.004)	**0.711**	/	**/**
Anti-osteoporosis therapy				
No	Ref	**Ref**	Ref	**Ref**
Yes	0.905 (0.541–1.513)	**0.705**	/	**/**
Multiple vertebral fracture				
No	Ref	**Ref**	Ref	**Ref**
Yes	2.638 (1.581–4.399)	**0.000**	2.456 (1.460–4.129)	**0.000**
Steroid use				
No	Ref	**Ref**	Ref	**Ref**
Yes	0.553 (0.261–1.173)	**0.122**	/	/

Relative Importance of Variables in Machine Learning Algorithms

The feature importance were different from each other while multiple vertebral fractures and bone density were at first. Intraoperative injury, length of surgery, height, age, body mass index and length of hospital stay were arranged in order of importance. Then, variables such as steroid hormone use, length of hospital stay to surgery, volume of bone cement injection, body weight, anti-osteoporosis medication, and gender had bad performance in predicting the contribution of bone cement leakage and were ranked in descending order (**Figure 3**). Particularly, re-fracture had no effect on bone cement leakage.

Web-Based Calculator

In this study, a web-based online calculator based on random forest model was developed. By entering variables that characterize the clinical profile of patients about to undergo percutaneous vertebroplasty, clinicians were able to predict the risk of cement leakage of their patients (https://share.streamlit.io/liuwencai0/pvp_leakage/main/pvp_leakage) (**Figure 4**).

DISCUSSION

Our research work used clinical data from patients to produce a risk calculator for bone cement leakage using ML algorithm, which using multiple variables to predict the risk oucomes of patients. The model was established based on a range of predictions and tested for accuracy and reasonableness through 10-fold cross validation. This artificial intelligence-based strategy can be exploited by clinicians to help them select more rational treatment responses (9).

ML contributes to the paradigm shift inherent in healthcare, where computers learn from patient data without being explicitly programmed to do the task (10, 11). It has the advantage of being highly capable, objective and repeatable in processing large data sets, reliable data (12, 13). It also has the potential to improve the quality of early diagnosis, identify disease progression and can increase the likelihood of predicting specific patient outcomes in orthopaedics, such as outcome scores, risk of complications and survival of implants (14). These benefits can facilitate the sharing of decision-making information between clinicians and patients and promote the effective planning and rational use of healthcare services (15, 16). In addition, the model can be retrained over time in order to improve predictive accuracy (17).

Percutaneous cement augmentation has become a common treatment for osteoporotic fractures of the thoracolumbar spine. Bone cement filling is an essential component that reconstructs the strength and biomechanics of the fractured vertebra and prevents re-fracture (18, 19). The incidence of bone cement leakage ranges from 5 to 87% (20, 21) and pulmonary embolism or spinal injury due to bone cement leakage is rarely reported (22, 23). Bone mineral density was an important indicator of bone strength and was the most effective predictor of fracture risk (24). Guo et al., found that the volume of injected bone cement was an independent risk factor for cement leakage (25), so we added this variable to our model algorithm.

In this study, we compared multiple machine learning algorithms for predicting the risk of bone cement leakage in patients undergoing percutaneous vertebroplasty. The comparison of ML algorithms revealed that the random forest performed best. To better utilize the model, we further built a web calculator estimating the probability of bone cement leakage in patients.

It was important to note that preoperative variables including age, body mass index and bone mineral density were significant predictors that can be used to assess the risk of cement leakage and thus help clinicians to choose the best treatment and know the prognosis of the patient. However, these factors in previous studies were based on preoperative information and were deficient to achieve reliable predictions. In addition, few studies have assessed the predictive value of intraoperative factors. Therefore, we established the key variable of intraoperative cement injection volume, which not only ensures that correct intraoperative

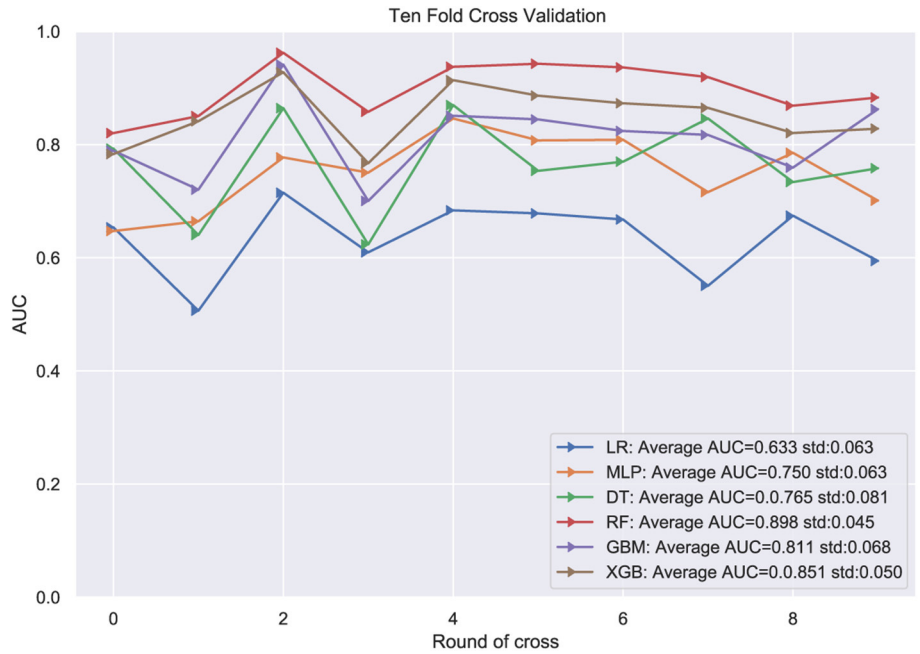


FIGURE 2 | 10-fold cross validation test. LR, logistic regression; MLP, Multilayer perceptron; DT, Decision Tree; RF, Random Forest; GBM, gradient boosting machine; XGB, extreme gradient boosting.

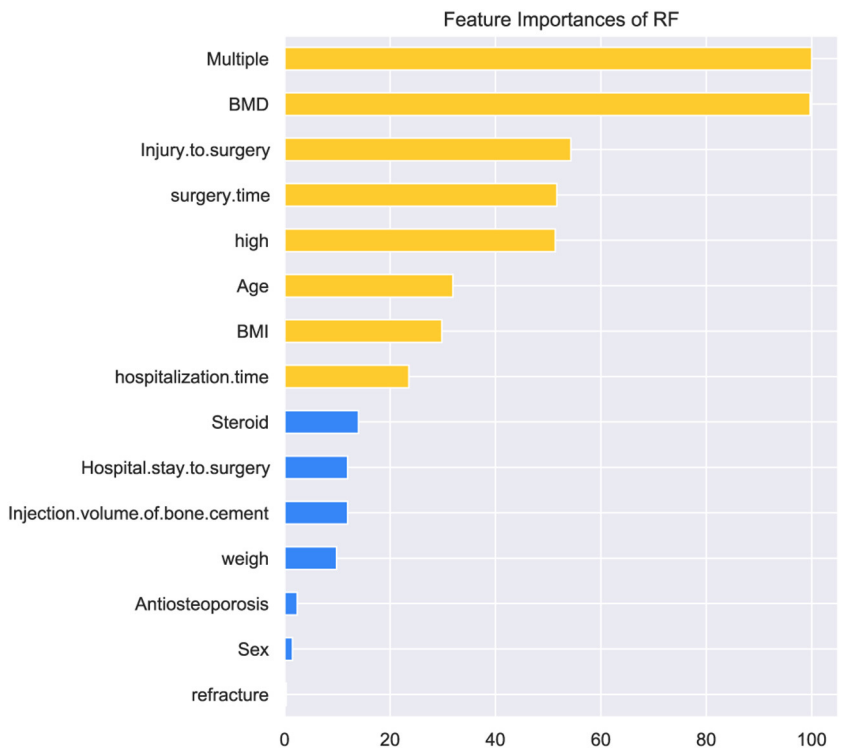


FIGURE 3 | Patient clinical feature importance of Random Forest.

manipulation and restoration of vertebral strength can be safely performed, but also helps us to predict the extent of cement leakage.

This study aimed to use a dataset consisting of 385 patients to help us select the best machine learning prediction. Our work had several advantages. First, few studies have used

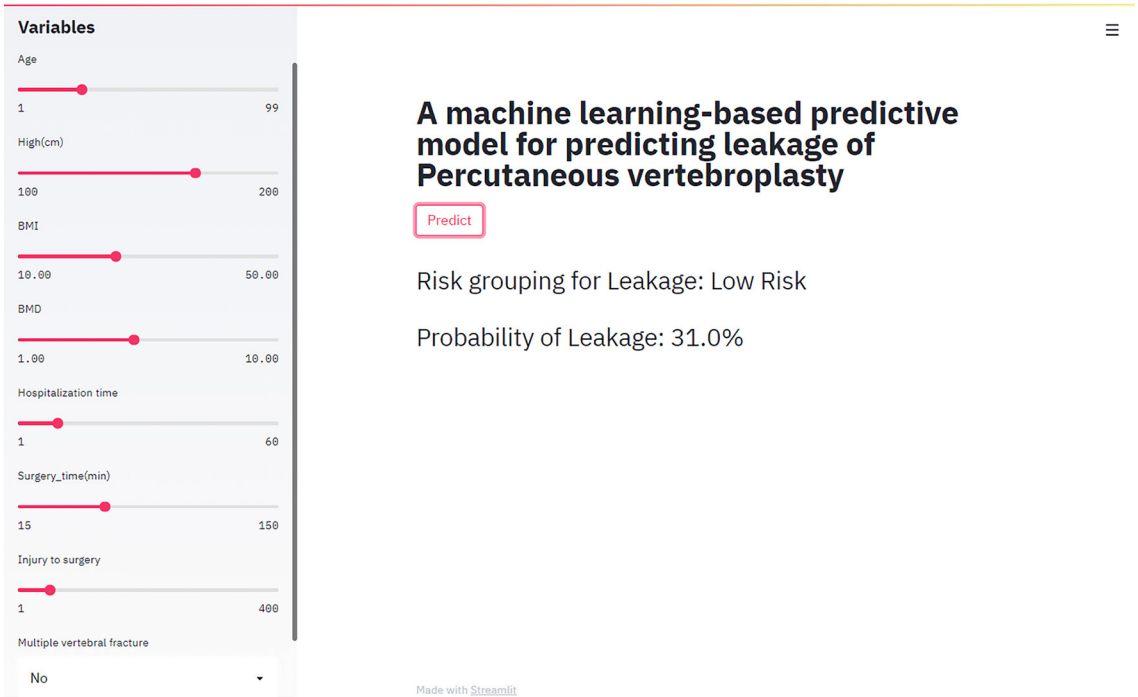


FIGURE 4 | The web-based calculator for predicting bone cement leakage in patients with percutaneous vertebroplasty.

machine learning methods to assess the risk of bone cement leakage during percutaneous vertebroplasty. Besides there were few studies in the existing literature that have applied ML algorithms to orthopaedic surgical complications. As far as we know, this study was the first to use ML algorithms to develop a web calculator that allows orthopaedic surgeons to use clinical data for real-time risk assessment of surgical complications. Last but not least, by comparison, our model shows superior predictive power and based on this we have built an online risk assessment program that allows orthopaedic surgeons to calculate risk variables and provide individualized surgical treatment to patients.

However, there are some limitations in this study. Firstly, the nature of retrospective studies can lead to subjective bias and selection bias. Secondly, ML algorithm model we developed was limited to one hospital. This may limit its use in other regions, which requires more use for further validation. Thirdly, the predictive performance was not good enough.

In machine learning, supervised learning identified the relationship between input data (e.g., patient clinical characteristics) and output data (e.g., bone cement leakage), which can help us predict outcomes. This has done through regression or classification. In classification tasks, the analysis was usually implemented by using logistic regression or decision trees. There were several studies on the latter shown that patient's age, associated comorbidities, functional status, etc. could lead the patient not being discharged as planned (26, 27). Furthermore, as more data being obtained, models can be retrained over time to improve prediction accuracy. Of course, clinicians always have the final decision when it comes to interpretation based on their domain expertise.

In conclusion, in this article we found that the random forest algorithm had good performance while applied as a tool for predicting bone cement leakage after orthopaedic surgery, Meanwhile, it was reasonably accurate and simple to use. Using this web-based calculator we could effectively prevent complications after percutaneous vertebroplasty and assist doctors in their surgical decisions. It is hoped that in the near future this web calculator will be able to cover a wide range of clinical variables, even at the molecular and cellular level, so that it can be used more accurately to assist a wider population.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This study data was approved by the Institutional Review Board of the Liuzhou People's Hospital before data collection and analysis, the IRB approval number is 2020 (KY-E-22-01).

AUTHOR CONTRIBUTIONS

CY and RL: designed the study. WLL, JW, and WLiu: performed the study and analyzed the data. WLL and JW: wrote the manuscript. QL and ZH: provided the expert consultations and clinical suggestions. CX and WYL: conceived of the study, participated in its design, and coordination. KZ, SS, and RL: helped to draft the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Lin EP, Ekholm S, Hiwatashi A, Westesson PL. Vertebroplasty: cement leakage into the disc increases the risk of new fracture of adjacent vertebral body. *Am J Neuroradiol.* (2004) 25:175–80. doi: 10.1055/s-2003-812630
2. Yoo KY, Jeong SW, Yoon W, Lee JU. Acute respiratory distress syndrome associated with pulmonary cement embolism following percutaneous vertebroplasty with polymethylmethacrylate. *Spine.* (2004) 29:294–7. doi: 10.1097/01.BRS.0000131211.87594.B0
3. Wu EQ, Hu D, Deng PY, Tang Z, Ren H. Nonparametric Bayesian Prior Inducing Deep Network for Automatic Detection of Cognitive Status. *IEEE Trans Cybern.* (2020) 51:1–14. doi: 10.1109/TCYB.2020.2977267
4. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-sanitization: a semantic privacy-preserving framework for unstructured medical datasets. *Comput Commun.* (2020) 161:160–71. doi: 10.1016/j.comcom.2020.07.032
5. Wu Y, Rao K, Liu J, Han C, Xu X. Machine learning algorithms for the prediction of central lymph node metastasis in patients with papillary thyroid cancer. *Front Endocrinol.* (2020) 11:577537. doi: 10.3389/fendo.2020.577537
6. Iwendi C, Bashir AK, Pasupuleti NS, Sujatha R, Jo O. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
7. Chen Z, Wu Y, Ning S, Ma T, Wu Z. Risk factors of secondary vertebral compression fracture after percutaneous vertebroplasty or kyphoplasty: a retrospective study of 650 patients. *Med Sci Monit.* (2019) 25:9255–61. doi: 10.12659/MSM.915312
8. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Mckinney EF. From big data to precision medicine. *Front Med.* (2019) 6:34. doi: 10.3389/fmed.2019.00034
9. Tang Z, Chen Y, Ye S, Hu R, Chang S. Fully memristive spiking-neuron learning framework and its applications on pattern recognition and edge detection. *Neurocomputing.* (2020) 403:80–7. doi: 10.1016/j.neucom.2020.04.012
10. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol.* (2018) 2:35. doi: 10.1186/s41747-018-0061-6
11. Iwendi C, Allen A. Enhanced security technique for wireless sensor network nodes. *IET Conf Wireless Sensor Systems.* (2012). 1–5. doi: 10.1049/cp.2012.0610
12. Devries Z, Hoda M, Rivers CS, Maher A, Phan P. Development of an unsupervised machine learning algorithm for the prognostication of walking ability in spinal cord injury patients. *Spine J.* (2019) 20:213–24. doi: 10.1016/j.spinee.2019.09.007
13. Bien N, Rajpurkar P, Ball RL, Irvin J, Lungren MP. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med.* (2018) 15:e1002699. doi: 10.1371/journal.pmed.1002699
14. Wu EQ, Deng PY, Qu XY, Tang Z, Sheng RSF. Detecting fatigue status of pilots based on deep learning network using eeg signals. *IEEE Trans Cogn Dev Syst.* (2020). 13:575–585. doi: 10.1109/TCDS.2019.2963476
15. Anajemba JH, Iwendi C, Mittal M, Tang Y, editors. Improved advance encryption standard with a privacy database structure for IoT nodes. *2020 IEEE 9th Int Conf Commun Syst Netw Technol.* (2020). 13:575–585. doi: 10.1109/CSNT48778.2020.9115741
16. Tang Z, Zhu R, Lin P, He J, Wang H, Huang Q, et al. A hardware friendly unsupervised memristive neural network with weight sharing mechanism. *Neurocomputing.* (2019) 332:193–202. doi: 10.1016/j.neucom.2018.12.049
17. Tang Z, Zhu R, Hu R, Chen Y, Chang S. A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors. *IEEE Trans Cogn Dev Syst.* (2020). 13:645–56. doi: 10.1109/TCDS.2020.3003377
18. Nam HGW, Jeong JH, Shin IY, Moon SM, Hwang HS. Clinical effects and radiological results of vertebroplasty: over a 2-year follow-up period. *Korean J Spine.* (2012) 9:334–9. doi: 10.14245/kjs.2012.9.4.334
19. Boger A, Wheeler KD, Schenk B, Heini PF. Clinical investigations of polymethylmethacrylate cement viscosity during vertebroplasty and related in vitro measurements. *Eur Spine J.* (2009) 18:1272–8. doi: 10.1007/s00586-009-1037-2
20. Nieuwenhuijse MJ, Erkel A, Dijkstra P. Cement leakage in percutaneous vertebroplasty for osteoporotic vertebral compression fractures: identification of risk factors. *Spine J.* (2011) 11:839–48. doi: 10.1016/j.spinee.2011.07.027
21. Zapalowicz K, Radek M. Percutaneous balloon kyphoplasty in the treatment of painful vertebral compression fractures: effect on local kyphosis and one-year outcomes in pain and disability. *Neurol Neurochir Pol.* (2015) 49:11–5. doi: 10.1016/j.pjnns.2014.11.005
22. Mpotsaris A, Abdolvahabi R, Hoffleith B, Nickel J, Weber W. Percutaneous vertebroplasty in vertebral compression fractures of benign or malignant origin: a prospective study of 1188 patients with follow-up of 12 months. *Dtsch Arzteblt Int.* (2011) 108:331–8. doi: 10.3238/arztebl.2011.0331
23. Vertebroplasty versus conservative treatment in acute osteoporotic vertebral compression fractures (Vertos II): an open-label randomised trial. Klazen CA, Lohle PN, de Vries J, et al. Lancet 2010;376(9746):1085–92. Epub 2010 Aug 9. *Spine J.* (2011) 11:88. doi: 10.1016/j.spinee.2010.11.011
24. Clung M. Romosozumab in postmenopausal women with low bone mineral density. *Dig World Core Med J.* (2014) 354:821—31. doi: 10.1177/0004563214528448
25. Guo R, Zhang K, Wen H, Zheng BL, Yang XB, Kong LB, et al. Risk factors for bone cement leakage in vertebral augmentation. *Orthop J China.* (2019). 40:1205–1210. doi: 10.1007/s00264-015-3102-2
26. Li W, Wang H, Dong S, Tang Z-R, Chen L, Cai X. Establishment and validation of a nomogram and web calculator for the risk of new vertebral compression fractures and cement leakage after percutaneous vertebroplasty in patients with osteoporotic vertebral compression fractures. *European Spine J.* (2021) 30:1–14. doi: 10.1007/s00586-021-07064-z
27. Morcos MW, Jiang F, Mcintosh G, Ahn H, Weber M. Predictive factors for discharge destination following posterior lumbar spinal fusion: a Canadian spine outcome and research network (CSORN) study. *Glob Spine J.* (2018) 9:403–8. doi: 10.1177/2192568218797090

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Wang, Liu, Xu, Li, Zhang, Su, Li, Hu, Liu, Lu and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 17 December 2021
doi: 10.3389/fpubh.2021.795007

An Explainable Intelligence Driven Query Prioritization Using Balanced Decision Tree Approach for Multi-Level Psychological Disorders Assessment

Sushruta Mishra[1], *Hrudaya Kumar Tripathy*[1], *Hiren Kumar Thakkar*[2*], *Deepak Garg*[3], *Ketan Kotecha*[4*] and *Sharnil Pandya*[5]

[1] School of Computer Engineering, Kalinga Institute of Industrial Technology, Deemed to Be University, Bhubaneswar, India, [2] Department of Computer Engineering, Marwadi Unversity, Rajkot, India, [3] Department of Computer Science and Engineering, School of Engineering and Sciences, Bennett University, Greater Noida, India, [4] Symbiosis Centre for Applied Artificial Intelligence, Symbosis International (Deemed) University, Pune, India, [5] Symbiosis Institute of Technology, Symbosis International (Deemed) University, Pune, India

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Neelakantam Gone,
Chang Gung University, Taiwan
Chintan Thacker,
Parul University, India
K. Gokulnath,
VIT-AP University, India
Abdul Rehman Javed,
Air University, Pakistan

***Correspondence:**
Hiren Kumar Thakkar
iamhiren@gmail.com
Ketan Kotecha
drketankotecha@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 14 October 2021
Accepted: 10 November 2021
Published: 17 December 2021

Citation:
Mishra S, Tripathy HK, Kumar Thakkar H, Garg D, Kotecha K and Pandya S (2021) An Explainable Intelligence Driven Query Prioritization Using Balanced Decision Tree Approach for Multi-Level Psychological Disorders Assessment. Front. Public Health 9:795007. doi: 10.3389/fpubh.2021.795007

Human emotions affect psychological health to a great level. Positive emotions relate to health improvement; whereas negative emotions may aggravate psychological disorders such as anxiety, stress, and depression. Although there exist several computational methods to predict psychological disorders, most of them provide a black-box view of uncertainty. This research involves developing a novel predictive model for multi class psychological risk recognition with an accurate explainable interface. Standard questionnaires are utilized as data set and a new approach called a Q-Prioritization is employed to drop insignificant questions from the data set. Moreover, a novel balanced decision tree method based on repetitive oversampling is applied for the training and testing of the model. Predictive nature along with its contributing factors are interpreted with three techniques such as permuted feature importance, contrastive explanation, and counterfactual method, which together form a reasoning engine. The prediction outcome generated an impressive performance with an aggregated accuracy of 98.25%. The mean precision, recall, and F-score metric recorded were 0.98, 0.977, and 0.979, respectively. Also, it was noted that without applying Q-Prioritization, the accuracy significantly drops to 90.25%. The error rate observed with our model was only 0.026. The proposed multi-level psychological disorder predictive model can successfully serve as an assistive deployment for medical experts in the effective treatment of mental health.

Keywords: psychological risks, explainable intelligence, decision tree, oversampling, predictive learning

1. INTRODUCTION

According to the observation of WHO, a healthy mind with physical well-being is possessed by a healthy individual (1). Also, it has been noted that various mental risks are gradually prevailing the society as a whole. In this fast-paced world with constant changes in the style of living, various kinds of psychological disorders are emerging in modern society. Mental

health concerns including anxiety, stress, and depression are common in people's life. Both personal and professional instabilities contribute to the occurrence of these disorders. Also, it is noted that such psychological concerns exhibit many overlapping symptoms. As an instance, individuals experience loneliness in all the three mentioned disorders (2). These health risks are vital parameters of psychological well-being and must be given proper attention at right time in order to prevent any potential negative impact on individuals (3). Usually, experts take the help of questionnaires and interactive sessions to assess these mental risks issues. As per their observation, psychologically affected patients reserve their sentiments within themselves and are reluctant to share them with family, friends, or even medical professionals. Common symptoms observed in an anxious person include nervous feeling, frequent irritation, insomnia accompanied by fatigue, panic tendency, fast breathing, rapid pulse rate, and concentration issues (4). Similarly being regularly upset, unable to relax, less energetic, having chronic headaches, and overreacting to situations are some primary symptoms noted in stress disorder (5). Similarly, depression exhibits some vital factors like lack of focus, memory fluctuations, poor decision making, lack of zeal in recreation, feeling of helplessness and restlessness, weight loss, and suicidal mindset (6). Thus, as it is discussed, many common overlapping symptoms are observed in these three psychological risks like pain in the chest, insomnia, fatigue, pulse rate rise, and lack of concentration. The presence of multiple overlapping symptoms makes effective categorization of these disorders a concern. The gradual rise in these psychological risks has led clinical researchers to focus their explorations in this domain. Accurate assessment, categorization, and treatment of anxiety, stress, and depression disorders are also a challenging task for machines (7). So a suitable learning model is needed for reliable assessment and diagnosis (8). Magnificent success in data analytics has led to numerous applications of artificial intelligence. Continuous rise in these techniques leads to developing intelligent models with self-ability to perceive, learn, and take a decision on its own (9). But the efficiency and robustness of these autonomous models are restricted by their incapability to interpret their outcomes and decisive functionalities to users. The real concern is on relying on decisions that lack proper justification, thereby unable to provide a detailed explanation of certain system behaviors (10). These predictive models provide a black-box view which is tough to explain. However, with the rapid rise in deploying these self-learning models in various application domains, the need for transparency and interpretability is highly required from all stakeholders in data analytics (11). Especially in critical health related applications it requires an extensive explanation of any decision to medical experts rather than a simple projection of output for diagnosis. Thus, in recent times with the modern adoption of complex learning models in the healthcare field, more attention is to develop human interpretability enabled frameworks that can enhance its implementation capability. Thus, a new approach called explainable intelligence needs to be adopted that are modeled to use predictive learning methods that perform the following functionalities.

- Generate explainable frameworks maintaining a high level of accuracy.
- Enabling end users to trust and understand the underlying functional characteristics.
- Ensuring impartial decision making.
- Determining all relevant factors which are essential for the prediction of an outcome.

Predictive models with such explainable intelligence capability can readily provide explanation toward any outcome, weighs pros and cons of a model and, thereby suggest on the future behavior of the developed model. These interpretability based frameworks can be integrated with system interfacing to transform a predictive decision into an upgraded reasoning prototype with relevant explanation to users.

The purpose of implementing explainable intelligent techniques is to appropriately explain the interpretation of predictive decisions obtained. Both the machine learning unit and the explainable intelligent unit are supplied with data and through the use of an explainable unit, precise explanations combined with prediction decisions are generated. These explanations can assist healthcare professionals to verify the decision of the predictive unit. Furthermore, it can be utilized with medical reports to provide in-depth reasoning and recommendations. A sample illustration of a predictive model supported by explainable intelligence is shown in **Figure 1**. Explainable intelligence is characterized by some vital elements which distinguish it from other conventional learning models. These critical elements concerned with an explainable intelligent model are discussed in this study.

- **Trust** is denoted by the certainty of a predictive model to perform during any troubleshooting.
- **Causality** is determined by the easiness to find associations such that there exists a strong logical link among variables.
- **Transferable** is governed by the interpretation of internal links occurring in a model which helps the user in reusing information to solve different problems.
- **Informative** property is based on the need for enormous information to reason predictive outcomes of a model, thereby avoiding local minima.
- **Fairness** symbolizes clarity of visualized results and the ethical ability to identify factors that affect a specific outcome.
- **Accessible** represents a feature allowing users to involve in the procedure of enhancing and building a predictive framework. Interactive is the capability of a learning model to share data and interact with the end user for effective interpretation.
- **Privacy** explores the functionality to interpret the internal links of a learning model by another third party to compromise the security aspects of the information.

2. PROBLEM STATEMENT

Psychological disorder risks such as anxiety, stress, and depression are quite common in modern society. There exist several prediction models that use different algorithms to detect

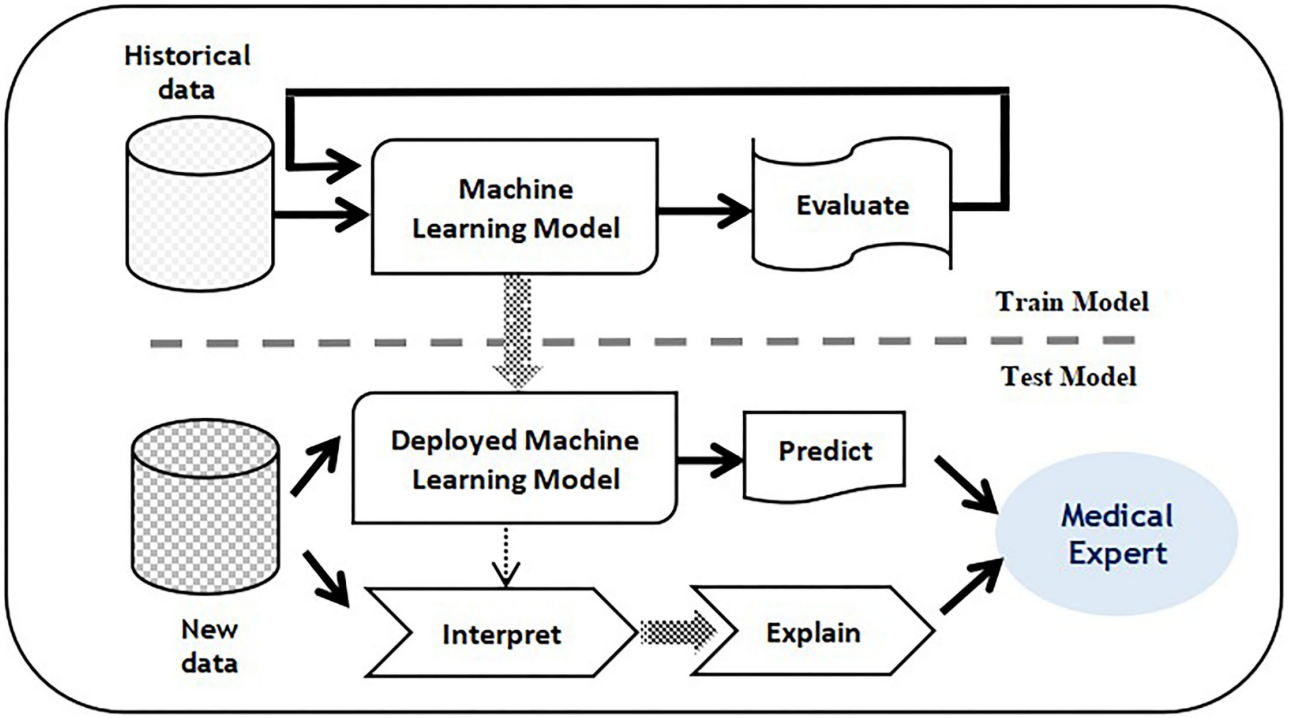


FIGURE 1 | Training and testing phase with proposed Balanced decision tree approach.

the presence of these mental health issues at an early stage. Unfortunately, these standard machine learning techniques present a black box view whose predictive nature cannot be easily interpreted by medical experts. With the rise of computationally intelligent models in the critical healthcare sector, effective reasoning and transparent functionalities are very essential since a contradictory predictive result may affect the mental stability of a patient. Therefore, a trade-off must be addressed between the prediction and transparent capability for critical use cases that almost benefit more reliably and implements a robust system to assist medical experts to find it easy to backtrack a certain psychological risk recognition. Thus, the high predictive ability should be well supported by an explainable interface to be more productive. In this study, we present an explainable intelligence-enabled psychological disorders predictive model comprising a transparent and reliable intelligent model to detect acute mental disorders from accumulated online questionnaires through interactive sessions with patients.

3. RELATED WORKS

Various academicians and researchers throughout the world have undertaken numerous quality oriented works related to healthcare prediction (12, 13) across different domains (14, 15) using recent technologies. Researchers have also analyzed psychological health disorders using many predictive learning models. Different datasets are used for this purpose. Evaluation of several existing models are also done to figure out the best performing models. But no model is singled out to deliver the optimum performance. Based on varying use cases, scenarios, dataset composition, training algorithm employed, and other miscellaneous factors, the performance fluctuates. Some high profile works which are significantly suited for the domain are discussed in this section. Dooshima et al. (16) have utilized learning models using several characteristic features such as biological, geographical, mental risks, and environment based variables to predict psychological disorders in patients. Many health professionals verified the authenticity of the model along with these metrics. Srividya et al. (17) framed a questionnaire to retrieve data of various features which was later used to detect mental risks in people. Among machine learning algorithms used in the study, nearest neighbor and random forest gave identical and good results. Dabek and Caban (7) applied an artificial neural network model for forecasting mental health ambiguities which include anxiousness and depression tendency. The study also stressed the impact of injuries and concussions on the health of sportspersons. A detailed analysis concerning various predictive models such as support vector machine, neural network, and k-nearest neighbor was undertaken for psychological patients by Alonso et al. (18). The review concluded that using information mining methods on depression and dementia-like disorders can yield fruitful outcomes and can enhance the life style of patients (19). Sau and Bhakta (20) applied several data analytics algorithms such as regression, boosting, and Naive Bayes to classify patients affected with mental disorders. In this study, around 500 data samples were collected through interviews,

and it was observed that boosting algorithm gave an optimal precision and accuracy value of 84.1 and 82.6%, respectively. A joint modeling architecture was developed by Saha et al. (21) where many linguistic variables on online sites were selected to classify psychological issues of virtual communities. The model performed better than other task learners, and the result explored the presence of sentiments beyond depression in the majority of online patients. Reece et al. (22) discussed the prediction parameters of depression and stress among the Twitter community where stress level was gauged through Hidden Markov Model (HMM). It was noted that around 24% and 32% of people were diagnosed with stress and depression tendency, respectively. Braithwaite et al. (23) used a decision tree model to determine patients with suicidal mindsets. About 135 tweets of different candidates were aggregated from Amazon Mechanical Turk and an accuracy rate of 92% was recorded. Du et al. (24) fetched Twitter samples with psychiatric stress units to tag suicidal tweets. Among all learning methods used, it was noted that convolution neural network performed the best among others including support vector machine and extra trees giving a 78% precision value in identifying online tweets with a suicide mindset. Al Hanai et al. (25) proposed a text-audio framework to detect depression where long short-term memory neural network was utilized to classify patients suffering from depression. The context-free method generated optimal results for multi-modal audio analysis. Ramiandrisoa et al. (26) predicted depression during early phases using social media data. CLEF eRisk tool was used to collect data. After an extensive evaluation, it was inferred that integration of information retrieval and machine learning produced the best outcome. Big data analytics was applied by Hou et al. (27) to forecast frustration and depression tendencies on basis of the reading pattern of users. Variables of Chinese textual data were retrieved to build a meta classification model. Naive Bayes gave the best performance among all. Leightley et al. (28) detected traumatic-related stress risks in ex-serviceman militants with supervised intelligent classifiers using various metrics such as gender, job state, and alcohol usage among others. It was observed that a good sensitivity rate was obtained for many classifiers except for the false negatives outcome. Young et al. (29) scanned facial expressions and body language of patients to detect mood variation and anxiety risks. The cross-validation method was used to generate a better precision rate and was alter validated by statistical algorithms. Mary et al. (30) conducted an experimental validation of stress, anxiety, and depression assessment with multiple machine intelligence methods using the DASS 21 dataset. Logistic regression gave the best performance with an accuracy of 90.33% on both stress and depression while 92% on anxiety risks. A detailed survey with around 6,000 candidates was done by Kessler et al. (31) to assess the degree of depression through the information mining approach and traditional methods. It was discovered that machine learners outperformed the traditional approach in detecting depression accurately. De Beurs et al. (32) presented an expert system with intervention mapping to help enhance user involvement in activities. The work discussed various patient specific design strategies for mental medicare using limited resources. Peng et al. (33) used kernel of support vector machine with some other learning classifiers to categorize depression patients on virtual community-based Twitter data. An accuracy rate of 83.46% was noted with kernel-based method. Anupriya et al. (34) analyzed various degrees of stress, depression, and anxiety risks using machine learning methods through DASS 2 questionnaires. Small dataset size was the major limitation of this study. Hatton et al. (35) assessed persistent depression in senior citizens through the gradient boosting method. In this study, PHQ-9 questionnaire was used for data collection purposes. Many such relevant existing research works are summarized in **Table 1**. Moreover, in Ray et al. (47), assessment of autistic disorder using machine learning approach is described. In Tripathy et al. (48), application and evaluation of classification model to detect autistic spectrum disorders in children are described. In Ray et al. (49), a review on Facial Expression Based Behavioral Analysis Using Computational Technique for Autistic Disorder Patients is described. In Mishra and Mohanty (50), an integration of machine learning and IoT for assisting medical experts in brain tumor diagnosis is described.

All these discussed work models provide good prediction and classification of psychological health risks. Various machine learning and attribute optimization methods are applied for precise analysis of such mental health concerns in humans. The majority of works used standard questionnaires based data set gathered through interactive face-to-face sessions. Also, the existing models have experimented on patients with varying age groups. Though all of these above mentioned works generate good performance, still they suffer from certain constraints. Problems related to the small sample size are a challenge for several deployed models. Another problem is the fact that data samples collected from interviews or manual queries based analysis restricts users from freely responding to any queries. As earlier observed, an individual having anxiety, stress, or depression symptoms are quite reserved within themselves rather than interacting with family, friends, or clinical staffs. They usually share their mind through anonymous platforms. But most importantly the biggest challenge lies in the fact that most of these predictive intelligent models simply provide black-box predictions without any reliable justified explanations of those mental health risk predictions. As a result, it creates ambiguities, and appropriate interpretations of a certain risk prediction can not be reasoned out. Thus, large sized dataset collected from online platforms are more desirable, and an explanatory interface is the need of the hour to properly interpret the psychological health prediction outcomes.

4. PROPOSED METHODOLOGY

The prime focus of the research is on recognizing physiological disorders such as anxiety, stress, and depression implementing an explainable intelligence based predictive model. Psychological disorder data samples in the form of responses of questionnaires are collected from the desired participants through Google forms and are stored for analysis. The proposed framework can be viewed to comprise certain interlinked functional phases which are discussed in this section.

TABLE 1 | Relevant works on psychological risk analysis using predictive analytics.

References	Reviewed risks	Approach	Inference
Bauer et al. (36)	Bipolar disorder	Paper-based survey	47% elderly people utilized internet and 87% youths exhibit bipolar disorder.
Dhaka and Johari (37)	Mental disorder	Genetic algorithm and MongoDB tool	Storage and processing massive mental risks data on MongoDB database.
Kumar and Bala (38)	Depression	Sentimental analysis and save data on Hadoop	Preprocessing online social media perspective on specific business products.
Furnham (39)	Personality disorder	Hogan "dark side" measure (HDS) concept of dependent personality disorder (DPD)	Most personality risk factors are highly linked to a type of cooperative personality.
Bleidorn and Hopwood (40)	Personality assessment	Prediction models and K-fold validation	Focused on aspects such as organized adaptability and arguments to improve verification of predictive techniques.
Sarraf and Tofighi (41)	Alzheimer's risk	Convolutional neural network	Mental health instances were successfully categorized with 96.86% accuracy rate.
Fiscon et al. (42)	Brain disorders	Decision tree and EEG signals	Decision tree outperforms others in precise risk detection with 90% accuracy and 87% specificity with use of cross validation method.
Chatterjee et al. (4)	Anxiety analysis	Regression and bayesian classifiers	Used a probabilistic technique to validate patients with anxiety levels. It concluded that Bayesian Network showed the best accuracy of 73.33%.
Omurca and Ekinci (43)	Traumatic stress risks	Neural networks and social media optimization	A hybrid system to classify PTSD individuals and allowed feature selection methods to find vital metrics of patients' risks. The accuracy differed between 74 and 79%.
Dabek and Caban (7)	Mental risks	Neural network	Analyzed 89,840 samples and recorded a classification accuracy of a range (73%-95%).
Katsis et al. (44)	Anxiety disorders	Integrated meta classifiers	Proposed a hybrid model with mental health signals for assessing anxiety risks. Accuracy of 77.33, 80.83, and 78.5% was the output with neural network, radial networks, and SVM, respectively.
Saxe et al. (45)	Stress risks	SVM and Lasso regression	Optimal AUC value noted was 79 and 78% with SVM and RF, respectively.
Karstoft et al. (46)	Stress and depression	Hybrid method Feature selection and SVM	Target Information Equivalence Algorithm optimized detection of PTSD when used with support vector machine. The mean AUC was 0.75.

1. Transform-Encode Phase.
2. Segregate-Label Phase.
3. Q-Prioritization Phase.
4. Train-Test Phase.
5. Prediction Interpret Phase.
6. Validate Phase.

The sequential six phases are required for a comprehensive psychological assessment. The Transform-Encode Phase is employed as the data samples are gathered in the form of the google form, the raw and unprocessed data need to be presented in a suitable template that can be easily interpreted. The Segregate-Label Phase is employed to partition the attributes of the data set into three distinct sets of physiological disorders such as anxiety, depression, and stress. After the data samples are pre-processed and labeled, it is followed by the Q-Prioritization phase which is responsible for dropping out less significant queries or attributes from the data set. Once the irrelevant attributes in the data set are eliminated, Train-Test Phase is used to learn through examples using an appropriate machine intelligence algorithm. The Prediction Interpret Phase provides a suitable explanation of the prediction outcomes.

4.1. Transform-Encode Phase

The retrieved data samples are gathered in the form of google responses from users. These raw unprocessed data need to be presented in a suitable template that can be easily interpreted. Thus, the collected responses are mapped onto a two-dimensional tabular structure by allowing it to be saved in a csv (comma separated values) file in a new editable excel sheet. If an online survey is followed then the Typeform tool can be applied to fetch the responses in the .csv file directly. If any other ways to fetch data in the spreadsheet is used then excel can be used in saving it as a .csv file. The collected data samples can be loaded and their attributes are summarized in the python platform with appropriate libraries in store. It is important to detect any missing values in the data as it affects the processing. Any data entries which are out of their predefined range are tagged as missing values. These values are counted, identified, and are accordingly replaced with the data value which is most common for that specific column. It is followed by the encoding of data where based on severity level, the responses are assigned numerical values in the range of 1 to 4. Finally, the rate point is determined on a [0,3] scale by subtracting 1 from the encoded value (51, 52).

4.2. Segregate-Label Phase

In this phase, the attributes of the dataset are partitioned into three distinct sets of physiological disorders which include anxiety, depression, and stress based on its belonging value defined at prior. Five severity levels of these disorders are considered and domain range for every level is set at prior corresponding to individual sets of anxiety, depression, and stress disorders. Normal (N), Mild (M), Moderate (MD), Severe (S), and Extremely Severe (ES) are the detected levels

TABLE 2 | A sample weight score computation illustration.

Q2	Q4	Q7	Q9	Q15	Q19	Q20	Q23	Q25	Q28	Q30	Q36	Q40	Q41	W_s	Class
Anxiety															
3	3	1	0	1	1	0	1	0	2	3	1	0	0	16	S
2	3	3	1	3	2	3	3	3	3	3	1	1	2	33	ES
3	0	0	0	0	0	1	0	2	0	1	2	0	0	9	M
0	2	0	0	0	0	1	0	0	0	1	1	0	1	6	N
1	0	2	1	0	1	2	0	1	0	0	1	2	0	11	MD
Q1	**Q6**	**Q8**	**Q11**	**Q12**	**Q14**	**Q18**	**Q22**	**Q27**	**Q29**	**Q32**	**Q33**	**Q35**	**Q39**	W_s	**Class**
Stress															
2	1	3	2	3	3	3	3	3	2	3	3	3	2	36	ES
3	2	2	3	3	1	1	3	0	1	1	0	1	2	23	MD
3	2	3	0	3	3	0	2	3	1	3	3	3	2	31	S
2	2	0	0	1	2	0	0	0	0	1	0	0	0	8	N
2	0	0	2	0	0	0	2	0	3	2	3	1	1	16	M
Q3	**Q5**	**Q10**	**Q13**	**Q16**	**Q17**	**Q21**	**Q24**	**Q26**	**Q31**	**Q34**	**Q37**	**Q38**	**Q42**	W_s	**Class**
Depression															
3	3	3	3	2	2	1	3	1	3	3	3	0	3	33	ES
0	1	2	1	0	2	1	1	0	0	0	2	1	0	11	M
1	0	0	0	1	0	0	0	0	1	0	0	0	1	4	N
1	2	1	0	0	0	3	3	1	3	0	2	3	3	22	S
3	1	2	0	0	1	0	1	2	2	2	1	1	2	18	MD

for the research work undertaken. The Weight Score (WS) for all disorder sets is calculated by summing up every data rate points for individual set of disorders taking into consideration of all column cell values and is given in Equation (1).

$$W_S = \sum RP^i \tag{1}$$

In this study, RP_i denotes the rate points of an individual set of mental disorders. "ES" label for anxiety, stress, and depression risks was assigned a WS of 20+, 33+, and 28+, respectively. "S" severity levels for these three classes were allotted a WS range of 15–19, 26–33, and 21–27, respectively. Similarly "MD" labeled instances were adjusted WS range of 10–14, 19–25, and 14–20, respectively. The WS values for the "M" class for all the risks were 8–9, 15–18, and 10–13, respectively. The WS values less than the above range are considered to be the "N" class.

A sample data-sheet illustrating the query set corresponding to anxiety, stress, and depression risks is represented in **Table 2**. It shows the computation of WS for all tuples and based on which the severity class label is assigned.

4.3. Q-Prioritization Phase

After the data samples are pre-processed and labeled, it is followed by the Q-Prioritization phase which is responsible for dropping out less significant queries or attributes from the dataset. It is to note that, the pre-processed question set is the input while a reduced and optimized question set is the output of the Q-prioritization phase. Sometimes it is observed that all attributes in the samples do not equally contribute to the processing and prediction process. The presence of these less relevant attributes affects the overall performance of the predictive model in the context of latency delay or accuracy of prediction. Thus, detecting and eliminating these less relevant features is a vital task. In this phase, a novel attribute relevance method is adopted before the model is subjected to training and testing using classifiers. The WS computed in the predecessor phase is analyzed and a relevant threshold (Rth) for every physiological disorder set is determined by Equation (2) as follows.

$$R_{th} = \frac{Normal_{max(PD)}}{Count_{col} \times Rate_{max}} \tag{2}$$

A R_{th} for every physiological disorder set is determined in Equation (2). In this study, $Normal_{max(PD)}$ represents the maximum value corresponding to the N level of any disorder set. The $Count_{col}$ denotes the number of attribute columns present in the questionnaire data set, and $Rate_{max}$ represents the optimum scaled rate point considered. Based on the equation, the R_{th} for every attribute column of the respective disorder set is computed.

The computed R_{th} value for anxiety, stress, and depression is 0.5, 1.0, and 0.2, respectively. Cumulative summation of all rate points which are grouped by individual column name is found out and a simple mean average for every column is determined. Furthermore, the priority of attribute columns is calculated on the basis of the ranking of their mean value in descending order which is calculated in terms of priority relevance (P^{rel}) as shown in Equation (3):

$$P^{rel} = \sum_{i=1}^{S} \frac{R_i}{S} \tag{3}$$

In this study P^{rel} is the priority relevance value to be computed. R_i denotes the rate point for every data sample for a specific disorder set, and "S" is the total samples of data collected. Based on the computation of P^{rel} obtained, those attributes are considered to be relevant and are retained. The overall pseudo code of this procedure is depicted in **Algorithm 1**.

Algorithm 1: Pseudocode for Q-prioritization phase.

Input: W_s
Output: Relevant prioritized queries.
Scan all labeled W_s determined in "Segregate-weight score phase" ;
Compute relevant threshold R_{th} for each *PD* subset: $R_{th} = \frac{Normal_{max(PD)}}{Count_{col} \times Rate_{max}}$;
foreach *disorder set D* **do** Add all rate points R_i grouped by column. ;
Find mean μ_i for each column. ;
Designate μ_i as P^{rel} ;
Rank all P^{rel} in ascending order priority: $P^{rel} = \sum_{i=1}^{S} \frac{R_i}{S}$;
if $P^{rel} \leq R_{th}$ **then**
  corresponding attribute column is irrelevant. ;
end
if $P^{rel} \geq R_{th}$ **then**
  corresponding column is relevant. ;
end
Relevant prioritized queries are retained. ;

4.4. Train-Test Phase

Once the irrelevant attributes in the dataset are eliminated, it is all set to learn through examples using an appropriate machine intelligence algorithm. After the model is adequately trained, it can be used to predict outcomes based on new data samples. In our research, an improved adaptation of the decision tree algorithm is deployed as the machine learning method. The decision tree is a widely used supervised approach to solve classification problems. It can be viewed as a hierarchical structure with internal nodes mapped onto attributes, branches denoting rules, and leaf nodes are the decision outputs. The decision node and leaf node are the two nodes of a decision tree. Decision nodes are responsible to take decisions while the outcome of decisions is denoted by leaf nodes. Based on the attributes of a dataset, decisions are made.

Algorithm 2: Decision tree building from training instances of data partition *P*.

Input: Data Partition *P*, Attribute_set, Attribute_select_method.
Output: Decision Tree.
Create node *s* ;
if *tuple* $\in$ *P with same label* **then**
  return *s* as leaf node with label *p*. ;
end
if *Attribute_set* == *NULL* **then**
  return *P* as a leaf node with major class in *s*. ;
end
Apply *Attribute_select_method*(*P*, *Attribute_set*) to check best split criteria. ;
Label node *s* with split_criteria. ;
if *split_attribute is discrete AND multi-way split permit* **then**
  Attribute_set = *Attribute_set* − *split_attribute* ;
  foreach *result I in split_criteria* **do** Let *P_i* be the tuple set in *P* satisfying result *i*. ;
  if P_i *is NULL* **then**
    A leaf attached to major class in *P* to node *S*. ;
  end
  else
    Attach node returned by *Build_Decision_tree*(P_i, *Attribute_set*)*tonodeS*. ;
  end
end
return *S*. ;

$$Gain(I, F) = Ent(I) - \sum_{s \in values(F)} \frac{|I_s|}{|I|} \times Ent(I_s) \tag{4}$$

A decision tree works with the purpose of predicting the label of a data sample initialized from the root node of the tree. Values of the root node are compared to the data variable and on basis of comparison follow the branch and move to the subsequent node. The process further compares the feature against all sub-nodes and proceeds. The same procedure is continued till the leaf node is reached. **Algorithm 2** relates its overall pseudocode.

One of the major concerns in a decision tree implementation is to choose the best attribute during the split. A suitable attribute selection measure solves the purpose. Information gain is a statistic based method for the selection of suitable attributes for testing at every node. It is a measure of entropy changes after attribute driven data segmentation which computes the information that an attribute gives regarding a class. Based on its value, the node is split, and accordingly, the decision tree is built. Gain (I; F) of an attribute A, that concerns a set of samples S is highlighted in Equation (4).

Where values (F) is the set of all possible values for attribute F, and Is is the subset of I for which the attribute F has value s.

Algorithm 3: Pseudocode for Information gain

Input: Instances, Attr, EntropyofSet
Output: Information Gain: $Gain(I, F)$

$Gain(I, F) = Ent(I)$;
for $Value \in Attr_Values(Instances, Attr)$ **do**
  $sub = subset(Instances, Attr, Value)$;
  $Gain(I, F) = Gain(I, F) - \frac{(count \in sub)}{number_of_samples} \times Ent(sub)$;
end
return $Gain(I, F)$;

The pseudocode for information gain is shown in **Algorithm 3**. Entropy in information theory is a metric that exhibits the purity of a random dataset. When the target feature accepts m distinct values, then its entropy I corresponding to m-wise categorization is denoted by Equation 5.

$$Ent(I) = \sum_{n=1}^{m} -P_i \times log_2 P_n \tag{5}$$

where "Ent" represents the entropy and pi denotes the probability of I which belongs to a class i.e., the logarithm of base 2 refers to the entropy of encoded size determined in bits.

4.4.1. Proposed Balanced Decision Tree Method

It is observed that the majority of predictive learning techniques generate optimistic classification performance which is misleading at times due to the uneven distribution of data samples in classes. It is due to the fact that many computational algorithms are built to function on classification samples with identical observations in every class. Data sampling can be used to deal with this issue a sit constitutes a set of methods to map an unbalanced data set onto more evenly distributed classes. The decision tree is biased toward the majority class, and as a result, the mis-classification rate of the minor class is more. In this study, sampling process can be fruitful in tackling this uneven data distribution. One such type called oversampling process replicates existing samples from minor classes and augment those to the train set. Here, training samples are randomly chosen with replacement such that a more balanced data set can be created. A balanced decision tree model differs from a simple decision tree in that it incorporates the oversampling technique to create an even distribution of data samples in both major and minor classes. The performance of a balanced decision tree is much better than the simple decision tree.

The improved decision tree presented in this study utilizes the oversampling method to generate a more efficient model. The physiological disorder data set is pre-processed and less relevant attributes are removed using the attribute relevance phase. The entire dataset is then partitioned into anxiety, stress, and depression sets on the basis of the predefined attribute set. The data instances are mapped onto the five disorder levels using simple if-then rules. Based on the cumulative rate point of a specific row, that instance falls in its corresponding disorder level. The labeling step is followed by oversampling where the minority class is identified in the training samples and random replicas of its instances are created. The sampling rate "S" chosen is 0.1. A decision tree algorithm is applied to the resultant oversampled dataset, and the accuracy rate is computed. The initial accuracy before the first iteration is assumed to be zero (Aold = 0). If the recent computed accuracy is found to be more than the predecessor value then the latest value is stored, thereby incrementing the sampling rate by 0.1. Again the data is oversampled with S = 0.1 to find the new accuracy. The process is repeated as long as the new accuracy Anew for a round is less than its previous value. At that point, the algorithm stops and generates the maximum accuracy value obtained throughout the process and is restored. After the training of the decision tree is done, the new test data is subjected to the model to predict the level of physiological disorder in the patient.

Thus, the developed model successfully detects the presence of any mental risk disorder, and if any inconsistencies are found then it predicts the level of disorder that the patient has probably developed corresponding to each set of a physiological disorder (anxiety, stress, and depression). Since the oversampling process is repeated multiple times in learning of decision tree to form a more evenly balanced data set, it is referred to as Balanced Decision Tree model which is illustrated in **Figure 2**.

4.5. Prediction Interpret Phase

An important observation that should be addressed is the set of attributes that have more impact on prediction outcomes. Most predictive models focus simply on the prediction performance rather than providing a suitable explanation for it. This creates an ambiguous scenario where factors that contributes to a certain prediction can not be determined. This creates a concern, especially when dealing with the real time critical healthcare domain. In our research analysis of detecting psychological health risk severity, an appropriate interpretation of predictive results is necessary from the medical experts and the patient's perspective. Thus, a predict interpret phase is augmented in our proposed model to facilitate the explainable intelligence functionality. This explainable intelligence capability is operated through a reasoning engine. The main aim of this reasoning engine is to allot a scoring value to all attributes on basis of their individual usefulness and relevance to a certain prediction. This module helps in providing a detailed insight in to the dataset. The scoring level of attributes can identify their relevance ranking which can be further used by medical experts to reason a particular mental health risk prediction. Our model implements a novel reasoning engine that consists of permuted feature importance, contrastive explanation, and counterfactuals methods for empowering the explainable intelligence functionality. The permutation feature importance explores the relationship between a specific attribute and the target label where a reduction in score indicates the dependency of that attribute on the developed model. The contrastive explanation method (CEM) aims to provide local interpretations to an abstract black box system that discusses

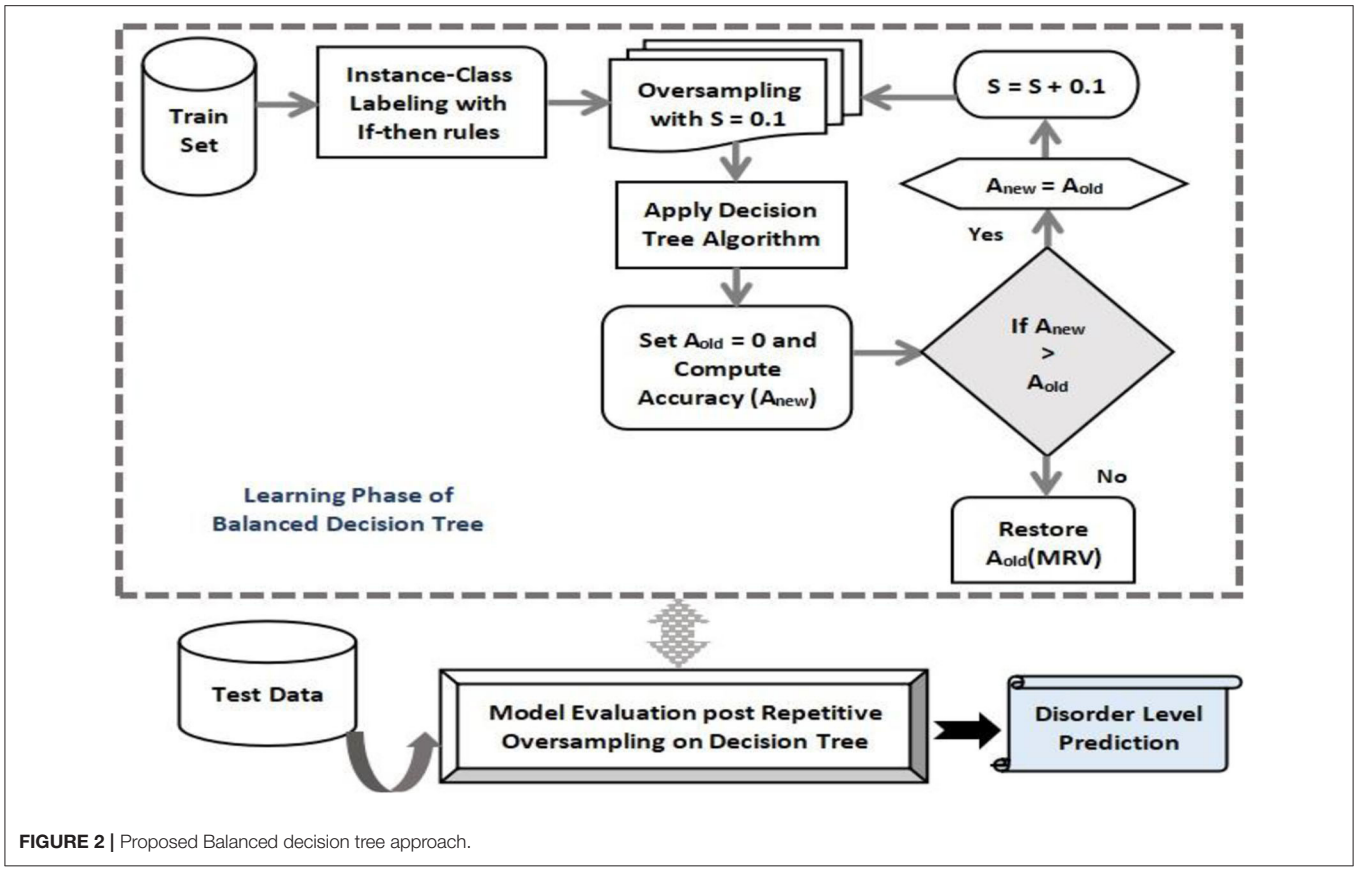


FIGURE 2 | Proposed Balanced decision tree approach.

explanation for classification by detecting less relevant features. A counterfactual interpretation of a prediction defines the variation in features that alters the prediction to a predetermined outcome.

4.5.1. Permuted Feature Importance Method

Permuted feature importance is an approach to calculate the relative significance scores of attributes and is independent of the model used. It acts as a non-local interpretation approach providing an in-depth idea of the nature of a predictive learner. Attributes are ranked on basis of the significance of every attribute on the trained learning algorithm predictive decision. Feature permutation importance computes the prediction ability of an attribute for a black box determinator by exploring the increase in prediction error if an attribute is absent. It begins with the calculation of base error on a trained predictive model. For each attribute, random shuffling of the corresponding column of trained data is performed, and the predicted error of the varied samples is computed. Then the feature importance is determined and stored as the difference between the base score and the varied sampled score. These steps are repeated multiple times to determine the average as it reduces the impact of random shuffling. Subsequently, on basis of the mean relevance of attributes on the score of the model, their ranking is done. Attributes with high score relevance on shuffling are labeled as more significant compared to attributes with less impact on the score of the model. Its pseudocode is shown in **Algorithm 4**.

Algorithm 4: Pseudocode for permuted feature importance method.

Input: Predicted model P, Dataset S
Output: Gain
Computer accuracy score c of the model P on dataset S. ;
for *each attribute column* $n \in S$ **do**
  for *each iteration* $m \in M$ **do**
    Randomly suffle n ;
    Generate referenced dataset $S_{m,n}$;
    Compute $c_{m,n}$ on dataset $S_{m,n}$;
  end
end
Computer importance $P - score$ for attribute A_n:
$P - score = c - \frac{1}{M} \sum_{m=1}^{M} c_{m,n}$;
Sort attributes in descending order of $P - score$;

Permutation feature selection can be used *via* the permutation_importance() function that takes a fit model, a dataset (train or test dataset is fine), and a scoring function P-score. The relative attribute importance can be illustrated in a bar chart where the x-axis shows the importance value and the y-axis denotes the attributes in significance order. A long bar is an indicator for an attribute with more relevance.

4.5.2. Contrastive Explanation Method

Contrastive explanation method is another explainable method that derives local interpretations of a black box architecture. Technically it may be defined as "An input m is categorized into class n since attributes ai,…aj are available and since attributes ax,…ay are unavailable." This approach suits more for tabular sample data which explains a classification model by giving details on relevant attributes or pertinent positives (PP) as well as irrelevant attributes or pertinent negatives (PN). It is a process that reasons out what may be minimally present and which variables may be essentially absent from the samples so that the original predicted class can be maintained. The two interpretation sides can be denoted as follows:

Pertinent Positives: An explanation to the PP figures out the attributes that are vital for model to predict the same output class as the predicted class. So from our research context, it consists of the crucial queries of the psychological risk data that impact most on the prediction outcome. The above result in **Figure 3** highlights the prediction of mental risk severity level is the same when PP is applied. The generated CEM values near 0 since those are the least values present to predict specific risk levels and these queries are to be essentially present to obtain the same original level as the severity level predicted S.

Pertinent Negatives: Explanation to PN computes the attributes which are to be adequately unavailable, form a sample, thereby retaining the original output level.

It can be seen from **Figure 4** that some of the CEM values in the array such as Q1, Q3, and Q5 are different from the original values, and these values alter the prediction class from "MD" to "S." Hence, we may conclude that variation in these attribute queries may essentially be absent in order to restore the original predicted class since they can flip the class label. Using CEM, we can improve the accuracy of the predictive model by analyzing the misclassified samples and subsequently processing those through CEM explanations.

4.5.3. Counterfactuals Method

Counterfactuals method provides a somehow "what-if" explanations based on prototypes where an instance represent the whole sample set of a class. Technically, it can be said that if M and N are independent and dependent variables, respectively, then counterfactual depicts the impact on N as a result of minor variation in M. Thus, it notifies the variation to be performed on M so as to alter the output from N to N. It is fast and gives quite accurate results by interpreting results with the use of prototypes of the target class. In context to our research analysis, a counterfactual sample denotes the vital modification in input query set of a testing sample which change the prediction to a predetermined outcome. A counterfactual sample Sxy should possess some basic characteristics which are outlined in this study.

1. The prediction over S_{xy} should be enclosed in a predetermined output label.
2. Variation β of the initial sample S_0 over $S_{xy} = S_0 + \beta$ needs to be sparse.
3. The counterfactual S_{xy} should be near to the general counterfactual class sample sharing.
4. The counterfactual S_{xy} should be computed fast to be applied in real time use cases.

These features can be approved by including a loss function in the form denoted by Equation 6 as:

$$Loss = xL_{Pred} + \beta L_1 + L_2 \tag{6}$$

Where xL_{pred} denotes the instance variation for predicting another class label than the original class. $\beta L_1 + L_2$ acts as the regularizer that defines sparsity by considering the net difference between counterfactual and alternate samples. In this study, as shown in **Figure 5**, samples of mentally disordered person with anxiety were collected and diagnosed. After diagnosis and an evaluation prototype was created to detect variations in various attributes to analyse risk possibilities. The observation highlights certain modifications to alter the outcome. The query Q1 (being upset) should be decreased by 0.76532 to avoid getting "S" risk level. Q12 (being nervous) should be under control by 0.182345 units. Similarly, as noted in Q29(–2.5674) and Q35(–0.5573), the patient should calm down more and be more tolerant to prevent being in "S" level of risk.

4.5.4. Validate Phase

Usually among all psychological health response data instances used as input, few among them may be misclassified to incorrect class in a predictive model. Thus, the performance of the proposed explainable intelligence enabled predictive model needs to be validated to figure out its effectiveness. Both training and testing of the model require validation using different metrics. Various performance parameters are involved in demonstrating its efficiency in detecting correct outcomes. Accuracy is a vital parameter that indicates the total mental health disorder instances that were correctly predicted among all the instances which were used as input to the predictive model. It is depicted in Equation 7 as:

$$Accuracy = \frac{TP + TN}{FP + FN} \tag{7}$$

Precision denotes the number of accurate mental health risk instances predicted which in reality maps to be positive. It is a fruitful indicator where false positives is a concern than false negatives. It is shown in Equation 8 as:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The recall represents the actual positive risk instances that accurately predict the model, and it is a very helpful indicator where false negative exceeds false positives. It is represented in Equation 9 as:

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

In reality, if the precision is increased then the recall value decreases and vice-versa. To capture both metrics simultaneously, another measure called F-score is used to present a combined concept of both. It denotes the harmonic average of

```
mode = ' PP'
lm = load_model(' stress.s5' )
cm = CEM(lr, mode, shape, kappa=kappa, beta=beta,
attr_range=attr_range, max_iterations=max_iterations,
c_init=c_init, c_steps=c_steps, learning_rate_init=lr_init,
clip=clip)
cm.fit(x_train, no_info_type=' median' )
interpret = cm.explain(X, verbose=False)
print(' Original sample: {}' .format(explanation.X))
print(' Prediction label: {}' .format([explanation.X_pred]))
print(' Pertinent positive: {}' .format(explanation.PP))
print(' Predicted label: {}' .format([explanation.PP_pred]))

.....................................................

Original sample: {{-1.0202; -1.342; -0.166; -0.785; -1.918}}
Prediction label: { 'S' }
Pertinent positive: {{-4.639e-09; -3.227e-03; -3.571e-05; -5.128e-4;
-4.952e-06}}
Predicted label: { 'S' }
```

FIGURE 3 | Pertinent positive sample example.

both recall and precision. The F-Score metric is denoted in the Equation 10 as:

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

Mean square error (MSE) is another relevant metric that accepts the mean value of the square of the difference of the original and the predictive values which are represented by Equation 11 as:

$$MSE = \frac{1}{M}\sum_{i=1}^{M}(Z_i - \bar{Z}_i)^2 \tag{11}$$

Where M represents the data points count, Z_i is the observed value, and $\bar{Z}_i$ is the predicted value.

Thus, as it is observed, all the phases of the proposed explainable intelligence model are interrelated such that the output of one phase acts as the input for the other. The overall working model is summarized in **Figure 6**. The psychological health risk data DASS 42 is retrieved online. The dataset is in the form of 42 questionnaires along with its responses. The original data samples are mapped onto an excel sheet for analysis. It undergoes appropriate pre-processing tasks like handling missing values and any inconsistencies in the dataset. The data is then encoded in the range of [1–4] based on the response of patients and scaled up to [0–3] for uniformity and ease of analysis. The dataset is grouped by predefined queries to partition into three subsets on the basis of psychological health disorders. Anxiety, stress, and depression are the identified subsets. The WS of all rows is determined, and accordingly, the risk severity level is labeled. The least priority queries are identified using the attribute mean rate point method and are eliminated from the dataset in Q-Prioritization phase. Then the resultant data is trained with a novel, newly proposed balanced decision tree technique which involves repeated oversampling. Once the model is trained, it can predict the appropriate disorder severity level for anxiety, stress, and depression risks with new unseen psychological data of new patients. The prediction outcome is further analyzed and interpreted using the reasoning engine which is comprised of three constituents. The first component called permuted feature importance method ranks all queries in priority order

```
n = 1
M = m_test[n].reshape((1,) + n_test[n].shape)
mode = ' PN'
shape = (1,) + n_train.shape[1:]
attr_range = (n_train.min(axis=0).reshape(shape)-.1,
n_train.max(axis=0).reshape(shape)+.1)
lm = load_model(' stress.s5' )
cm = CEM(lr, mode, shape, kappa=kappa, beta=beta,
attr_range=attr_range, max_iterations=max_iterations,
c_init=c_init, c_steps=c_steps, learning_rate_init=lr_init,
clip=clip)
cm.fit(x_train, no_info_type=' median' )
interpret = cm.explain(X, stress=False)
print(' Attribute names: {}' .format(attr_names))
print(' Original sample: {}' .format(explanation.M))
print(' Prediction label: {}' .format([explanation.M_pred]))
print(' Pertinent negative: {}' .format(explanation.PN))
print(' Predicted label: {}' .format([explanation.PN_pred]))

.....................................................

Attribute names: {Q1, Q12, Q22, Q29, Q35}
Original sample: {{-1.0202; -1.342; -0.166; -0.785; -1.918}}
Prediction label: { 'MD' }
Pertinent negative: {{-1.365; -1.342; -0.131; -0.785; -1.425}}
Predicted label: { 'S' }
```

FIGURE 4 | Pertinent negative sample example.

to identify the most relevant queries which contributed to prediction. The contrastive explanation approach further helps in recognizing the queries which should be present and similar queries that should be absent for the prediction of the severity risk level. Based on this, an interpreted tabular feedback is generated for reference of medical experts. The counterfactuals method is the third reasoning method that defines the required variations required in input queries in order to flip the generated severity class label. Thus, this reasoning engine forms the most vital part of the model a sit empowers the predictive model with an explainable intelligence capability.

```
Y = y_test[target].reshape((1,) + x_test[target].shape)
count = CounterFactualProto(nn, shape, use_ID3=True,
theta=10., max_iterations=1000,
attr_range=(x_train.min(axis=0), x_train.max(axis=0)),
c_init=1., c_steps=10)
count.fit(x_train)
interpret = count.explain(X)
print('Original prediction:
{}'.format(explanation.orig_class))
print('Counterfactual prediction:
{}'.format(explanation.count['class']))
on = X * sigma + mu
counterfact = explanation.count['X'] * sigma + mu
beta = counterfact - on
for j, l in enumerate(attr_names):
if p.abs(beta[0][j]) > 1e-4:
print('{}: {}'.format(f, beta[0][j]))

.....

Original predicted label: 'MD'
Counterfactual prediction: 'S'

Q1: -0.76532;
Q12: -18.2345;
Q22: 6.8002;
Q29: -2.5674;
Q35: -0.5573;
```

FIGURE 5 | Sample demonstration of counterfactual method.

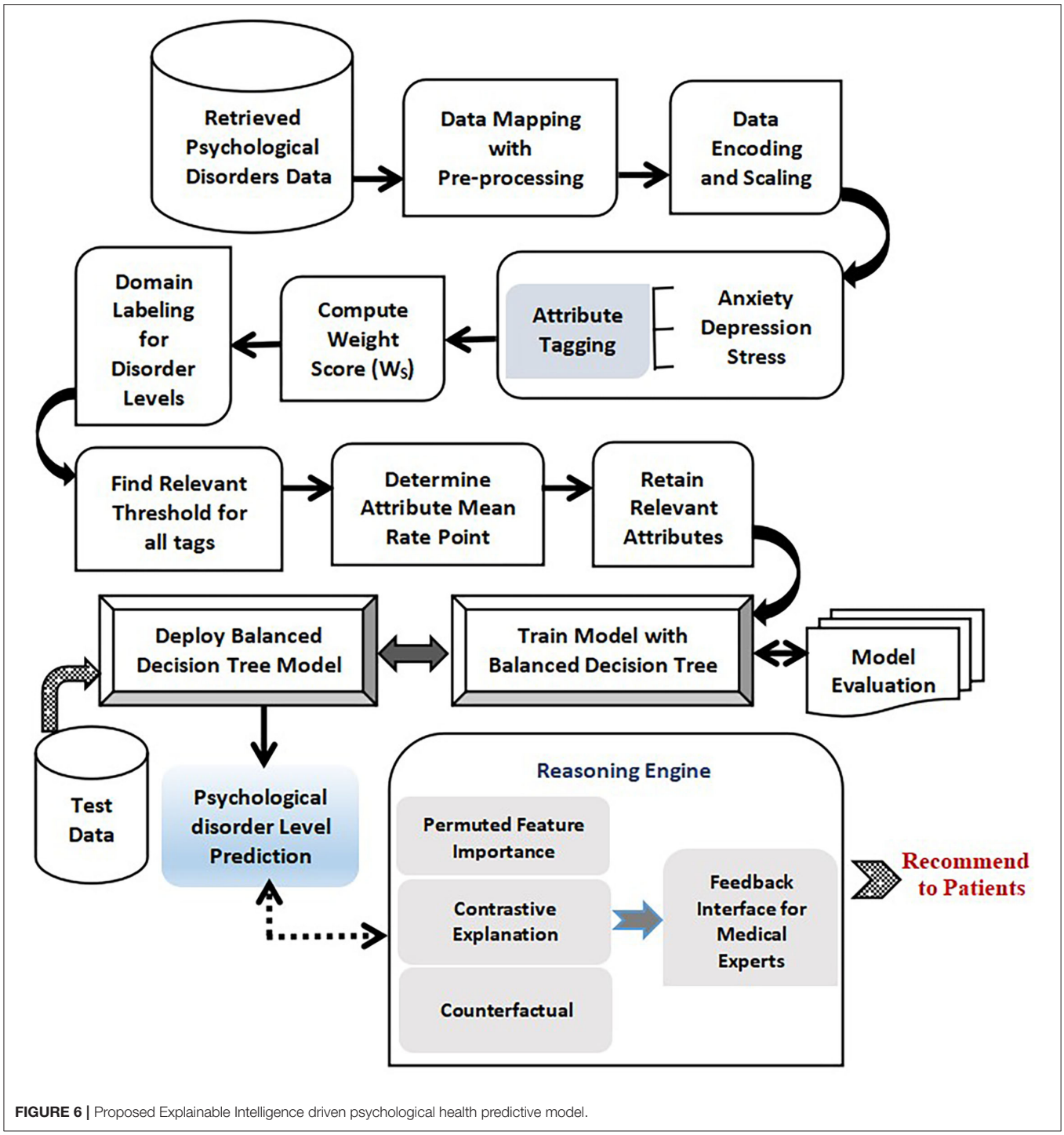


FIGURE 6 | Proposed Explainable Intelligence driven psychological health predictive model.

5. IMPLEMENTATION RESULTS AND ANALYSIS

The proposed explainable intelligence based mental health risk detection model is tested with the newly built balanced decision tree technique to predict the severity level of anxiety, stress, and depression health risks in patients. Before real time deployment, it is validated with different performance parameters to explore its efficiency in detecting psychological disorders.

5.1. Training Data Evaluation Performance

The psychological health disorder dataset used for performance evaluation is the standard DASS 42 questionnaires. As many as 13,000 responses of patients forms the dataset. With the help of the performance parameters indicated in the previous section, the

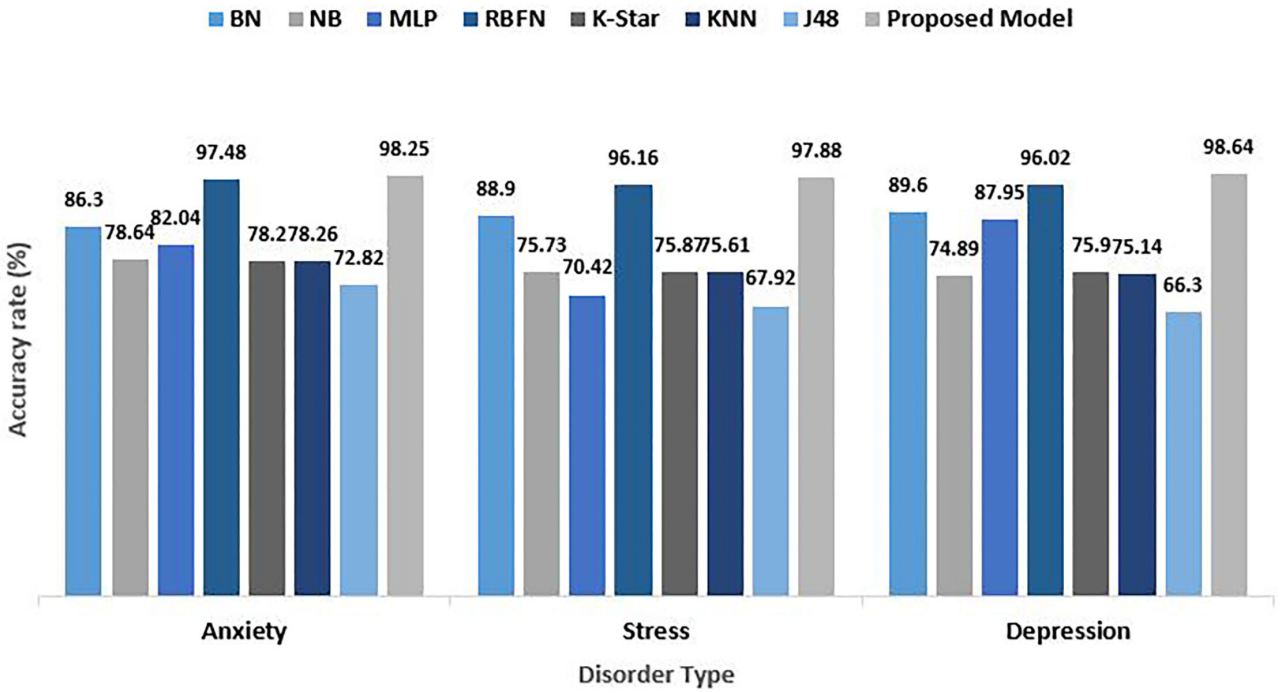


FIGURE 7 | Comparative analysis of accuracy rate for psychological risks.

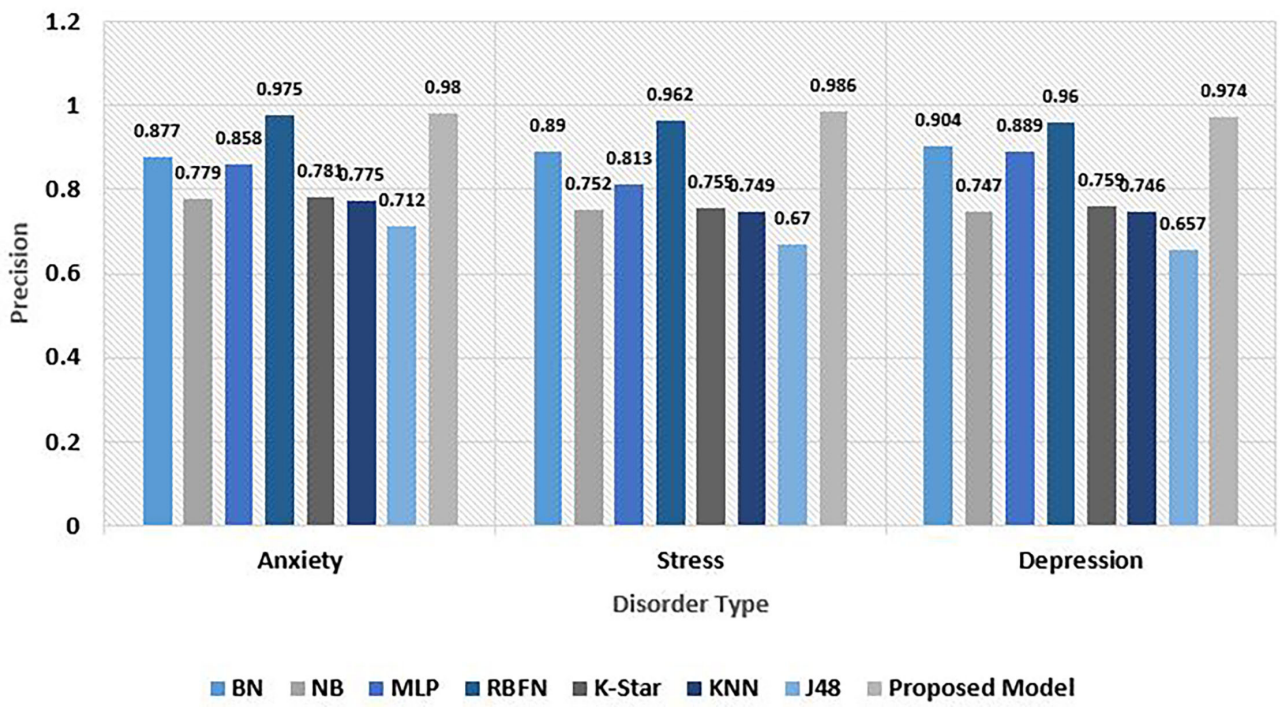


FIGURE 8 | Comparative analysis of precision for psychological risks.

performance is determined. The dataset is experimentally trained using many predictive models such as Naive Bayes, RBFN, KNN, J48, and neural network apart from the proposed model. Accuracy is the most important evaluation parameter which denotes the proportion of accurate predictions to the cumulative observations. Accuracy is a very relevant metric in an evenly

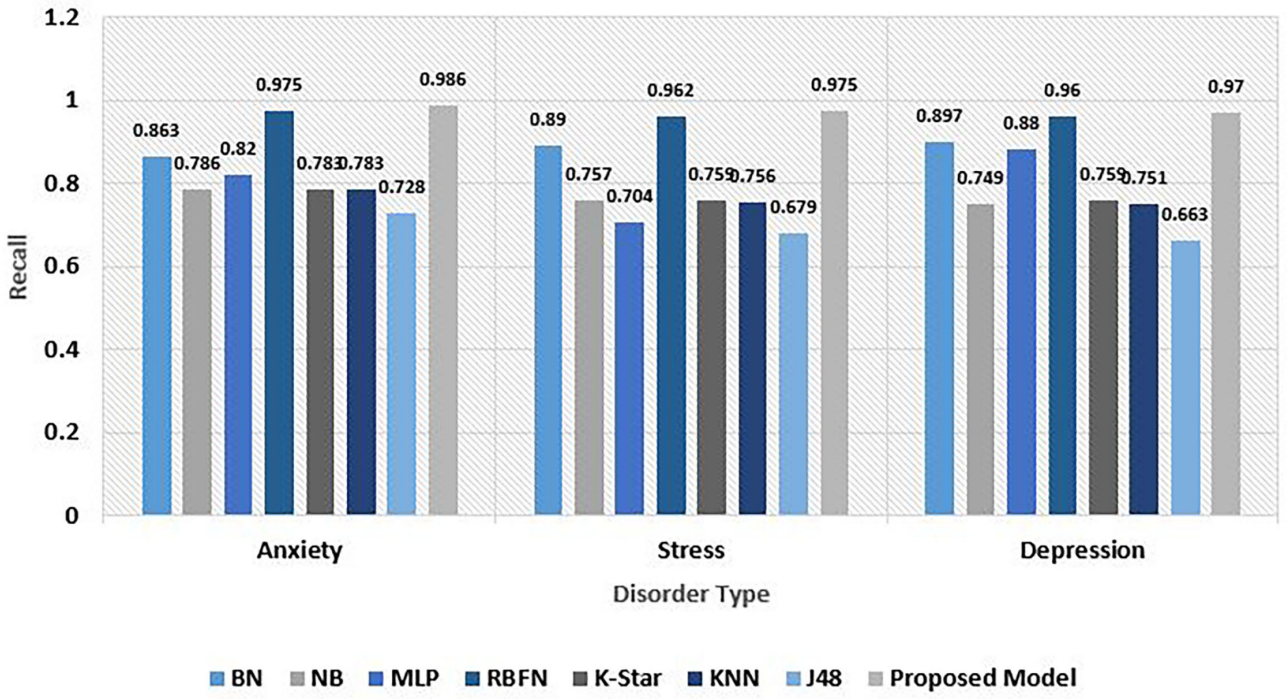


FIGURE 9 | Comparative analysis of Recall for psychological risks.

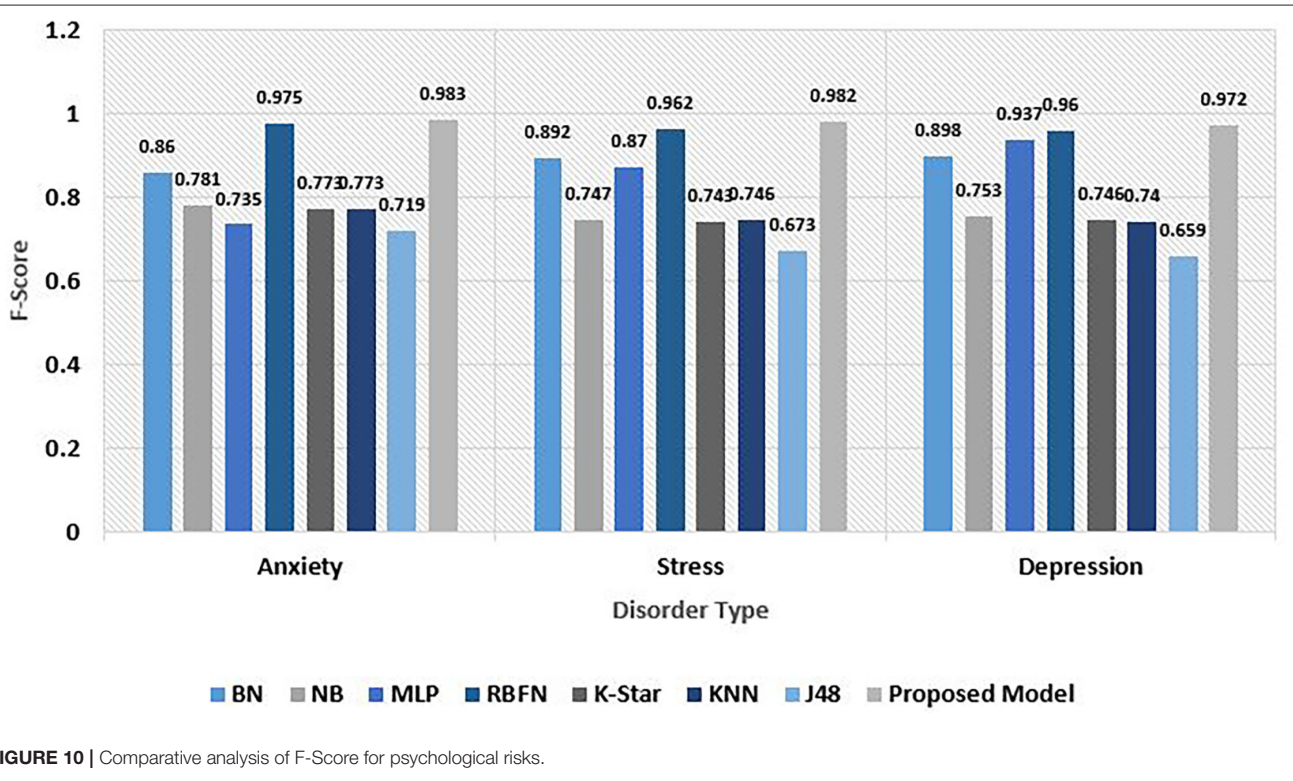


FIGURE 10 | Comparative analysis of F-Score for psychological risks.

balanced data samples. It is observed that among all existing techniques, the RBFN model provided an impressive accuracy on all three datasets while the accuracy of J48 was relatively discouraging. But our proposed explainable intelligence model outperformed others recording an accuracy rate of 98.25%, 97.88%, and 98.64% on anxiety, stress, and depression set,

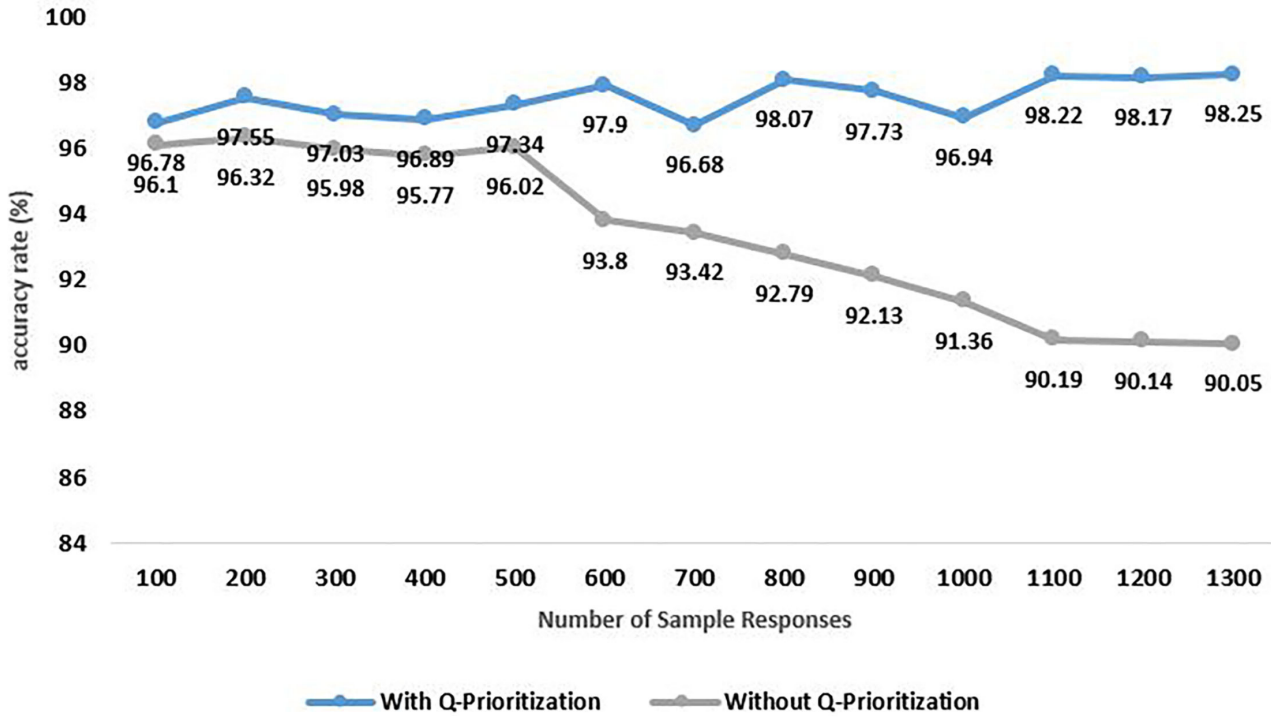


FIGURE 11 | Accuracy rate analysis of psychological risks in context to Q-Prioritization phase.

TABLE 3 | Performance metrics comparison using different datasets.

	DASS 42	DASS 21	
Accuracy	0.982	98.16	Anxiety
Precision	0.98	0.976	
Recall	0.986	0.982	
F-Score	0.983	0.978	
Accuracy	97.88%	97.69%	Stress
Precision	0.986	0.98	
Recall	0.975	0.968	
F-Score	0.982	0.975	
Accuracy	98.64%	98.42%	Depression
Precision	0.974	0.958	
Recall	0.97	0.962	
F-Score	0.972	0.96	

TABLE 4 | Queries associated with anxiety risk.

Query	Description
Q2	I was aware of dryness of my mouth.
Q4	I experienced breathing difficulty such as rapid breathing and breathlessness.
Q7	I had a feeling of shakiness such as legs going to give away.
Q9	I found myself in situations that made me so anxious I was most relieved when they ended.
Q15	I had a feeling of faintness.
Q19	I perspired noticeably in the absence of high temperatures of physical exertion.
Q20	I felt scared without any good reason.
Q23	I had difficulty in swallowing.
Q25	I was aware of the action of my heart in the absence of physical exertion.
Q28	I felt I was close to panic.
Q30	I feared that I would be thrown by some trivial but unfamiliar task.
Q36	I felt terrified.
Q40	I was worried about situation in which I might panic and make a fool of myself.
Q41	I experienced trembling in the hands.

respectively, as shown in **Figure 7**. The aggregated accuracy computed was 98.25%.

Precision represents the proportion of accurate positive prediction observations to the cumulative positive prediction based observations. Precision evaluates the accurately categorized samples against those samples classified to be positive. The model performed exceptionally well in terms of

TABLE 5 | Queries associated with stress risk.

Query	Description
Q1	I found myself getting upset by quite trivial things.
Q6	I tended to over-react to situations.
Q8	I found it difficult to relax.
Q11	I found myself getting upset rather easily.
Q12	I felt that I was using a lot of nervous energy.
Q14	I found myself getting impatient when I was delayed in any way.
Q18	I felt that I was rather touchy.
Q22	I had it hard to wind down.
Q27	I found that I was very irritable.
Q29	I found it hard to calm down after something upset me.
Q32	I found it difficult to tolerate interruptions to what I was doing.
Q33	I was in a state of nervous tension.
Q35	I was intolerant of anything that kept me from getting on with what I was doing.
Q39	I found myself getting agitated.

TABLE 6 | Queries associated with depression risk.

Query	Description
Q3	I could not seem to experience any positive feeling at all.
Q5	I just could not seem to get going.
Q10	I found that I had nothing to look forward to.
Q13	I feel sad and depressed.
Q16	I felt that I had lost interest in just about everything.
Q17	I felt I was not worth much as a person.
Q21	I felt that I was not worthwhile.
Q24	I could not seem to get any enjoyment out of the things I did.
Q26	I felt down-hearted and blue.
Q31	I was unable to become enthusiastic about anything.
Q34	I felt I was pretty worthless.
Q37	I could see nothing in the future to be hopeful about.
Q38	I felt that life was meaningless.
Q42	I found it difficult to work up the initiative to do things.

precision metrics on all three risks. A value of 0.98 was observed with anxiety data, 0.986 on stress data, and 0.974 on depression data. Thus, a simple mean value of 0.98 precision was noted leaving behind the RBFN model which has a 0.968 precision value. J48 performed relatively poor in comparison with others. **Figure 8** depicts the precision score analysis of the proposed explainable model with others.

Recall metric helps in the quantification of accurate positive predictions among all positive predictions performed. While precision gives an idea of accurate positive predictions, recall gives a measure of missed out positive predictions. For imbalanced learning, recall is typically used to measure the coverage of the minority class. The explainable driven predictive model outperforms all other existing models in terms of recall value on all disorder data. While an excellent value of 0.986 is the result with anxiety data, on stress and depression set, the recall metric evaluates to be 0.975 and 0.97, respectively. The aggregated recall observed was 0.977. The recall analysis is depicted in **Figure 9**.

F-score is a means to integrate both recall and precision metrics, thereby capturing both features. Once both of the metrics are computed for our multi-level classification, an F-score can be utilized to combine both generating a harmonic average value. F-score considers both false negatives and false positives. The F-score value recorded on the anxiety, stress, and depression risk set was 0.983, 0.982, and 0.972. The average mean F-score value generated was 0.979. Among other models, the RBFN algorithm recorded a very good value while J48 gave a comparatively inferior outcome. The overall F-score analysis was observed in **Figure 10**.

The impact of the Q-Prioritization phase on the performance was also explored and the outcomes are illustrated in **Figure 11**. The observation was quite clear. Right from the initial data size of 100 instances, the accuracy of the explainable model using Q-Prioritization phase, 96.78% was more than that of the model deployed without it, 96.1%. As we can see that the accuracy shows a significant drop once the dataset size exceeds 500 instances. The difference in accuracy increases with the rise in the size of data and toward the end of simulation with 1,300 samples. The final accuracy without the Q-Prioritization phase is found to be 90.05%.

The experimental evaluation was also undertaken with DASS 21 data set that constitutes 21 queries being mapped onto anxiety, stress, and depression risk. The same performance indicators were used for the evaluation. We obtained 98.09, 0.971, 0.97, and 0.972% as the mean accuracy, precision, recall, and F-score value obtained with DASS 21 sample set. This shows the consistent performance of our proposed model with heterogeneous data sets. The outcome summary is outlined in **Table 3**. There is variation in values because two different data sets (DASS 42 and DASS 21) are used for the evaluation of the proposed research on mental risk disorders. Both these data sets comprise dissimilar questions. Also, distinct questions are associated with both data sets so based on the feature importance of individual questions, the values for accuracy and other parameters are different.

5.2. Real Time Prediction Outcome

The experimental evaluation was undertaken in a real time environment. Our proposed model was validated with volunteers. DASS-42 is the standard dataset used for the purpose. The dataset constitutes a series of 42 distinct queries being asked to patients affected with psychological disorder risks. The queries were equally clustered into anxiety, stress, and depression which forms the three predefined sets as discussed earlier. So every risk set comprises 14 non-overlapping queries. **Tables 4–6** displays the 14 queries each chosen for anxiety, stress, and depression disorder set.

Figure 12 depicts the google form prepared for our research hypothesis. This form comprises a total of 42 distinct questionnaires which are mandatory to be filled up by users. This acts as a self-testing interface for the users in an attempt to test themselves of the presence of any psychological health risks and is called psychology parser. The user needs to enter his name and ID before appearing in the test.

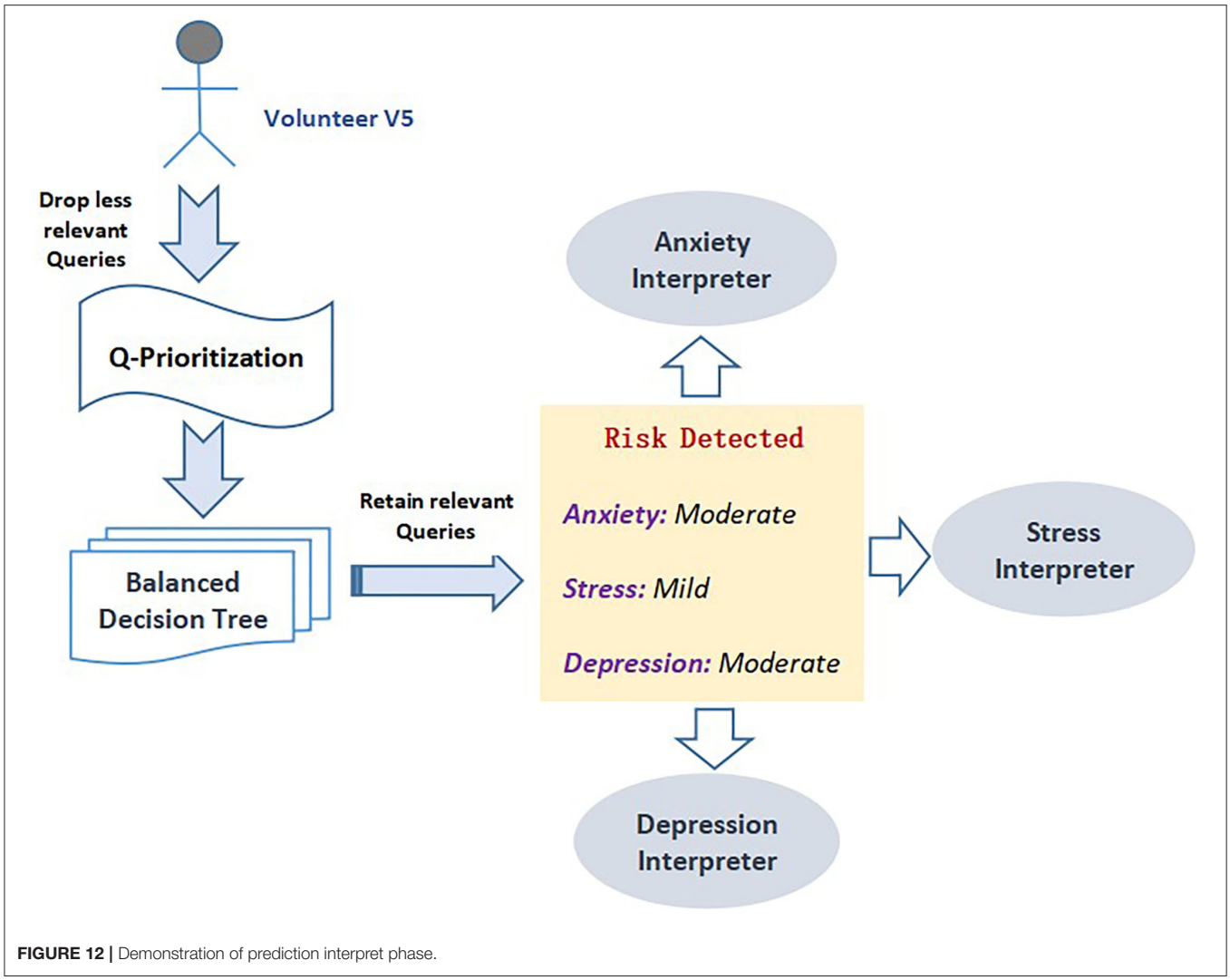


FIGURE 12 | Demonstration of prediction interpret phase.

The outcome demonstration for one specific day is discussed in this section where the test data of 10 volunteers are accumulated and their responses are mapped onto an excel sheet in a .csv file. It is followed by the important Q-Prioritization phase which is responsible for identifying less relevant queries from all sets and eliminating them using the mean Rate Point method. **Table 7** demonstrates this process. It starts with the original query set segregated into 14 queries in each disorder set including anxiety, stress, and depression. The individual queries are assigned their scaled-up value (0–3) based on the response of patients. The priority relevance value (Prel) for all queries (attributes) corresponding to individual disorder sets are determined using the mean rate point method.

This is followed by the ranking of the queries based on the P^{rel} value obtained. Accordingly, the least priority query that fails to satisfy the R_{th} is eliminated from the resultant data set. As it is observed in **Table 10**, query "Q9" was found to be less relevant in the anxiety set. Similarly, queries "Q18" and "Q5" in stress and depression set, respectively, were detected to be at less priority compared to others. Thus, these least priority queries were dropped from the data set after the Q-Prioritized phase. The least priority queries are marked in bold in **Table 8**.

The psychological parser is definitely technically feasible. It is a kind of google form software created at the back-end of model development in which all questions related to the research task is included and randomly shuffled. Patients are required to fill up the parser and then the risks of anxiety, stress, and depression are auto-computed. So it is completely software enabled with no complex hardware involved.

Once the data set undergoes the Q-Prioritized phase, it is pre-processed and is ready for prediction. Test data sample of any particular volunteer is subjected to the proposed Balanced Decision Tree algorithm model for detection of psychological disorder risks.

Figure 13 highlights the test data of volunteer "V5" as the input to the Q-Prioritization phase to find the priority based relevant query set. This query set is applied to the proposed balanced decision tree model. The prediction outcome I presented was in the form of an interface. If any risk is detected

TABLE 7 | Demonstration of Q-Prioritization phase.

	Q2	Q4	Q7	Q9	Q15	Q19	Q20	Q23	Q25	Q28	Q30	Q36	Q40	Q41
Anxiety risk set														
V1	3	3	1	0	1	1	0	1	0	2	3	1	0	0
V2	2	3	3	1	3	2	3	3	3	3	3	1	1	2
V3	3	0	0	0	0	0	1	0	2	0	1	2	0	0
V4	0	2	0	0	0	0	1	0	0	0	1	1	0	1
V5	1	0	2	1	0	1	2	0	1	0	0	1	2	0
V6	3	0	0	0	0	0	1	0	3	0	1	2	0	0
V7	0	2	0	0	0	0	1	0	0	0	0	1	0	1
V8	2	0	2	1	0	1	2	0	1	0	0	1	2	0
V9	3	3	1	0	1	2	0	1	0	2	3	1	0	0
V10	2	2	3	1	3	1	3	2	3	3	3	1	1	2
Mean	1.9	1.5	1.2	0.4	0.8	0.8	1.4	0.7	1.3	1.0	1.5	1.2	0.6	0.6
	Q1	**Q6**	**Q8**	**Q11**	**Q12**	**Q14**	**Q18**	**Q22**	**Q27**	**Q29**	**Q32**	**Q33**	**Q35**	**Q39**
Stress risk set														
V1	2	1	3	2	3	3	3	3	3	2	3	3	3	2
V2	3	2	2	3	3	1	1	3	0	1	1	0	1	2
V3	3	2	3	0	3	3	0	2	3	1	3	3	3	2
V4	2	2	0	0	1	2	0	0	0	0	1	0	0	0
V5	2	0	0	2	0	0	0	2	0	3	2	3	1	1
V6	2	3	0	0	1	2	0	0	0	0	1	0	0	1
V7	1	0	0	2	0	0	0	2	0	3	2	3	1	1
V8	3	2	3	0	3	3	0	2	3	1	3	3	3	1
V9	2	1	3	2	3	3	3	3	3	2	3	3	2	2
V10	3	2	2	3	3	1	1	3	0	1	1	0	1	1
Mean	2.3	1.5	1.6	1.4	2.0	1.8	0.8	2.0	1.2	1.4	2.0	1.8	1.5	1.3
	Q3	**Q5**	**Q10**	**Q13**	**Q16**	**Q17**	**Q21**	**Q24**	**Q26**	**Q31**	**Q34**	**Q37**	**Q38**	**Q42**
Depression Risk set														
V1	3	3	3	3	2	2	1	3	1	3	3	3	0	3
V2	0	1	2	1	0	2	1	1	0	0	0	2	1	0
V3	1	0	0	0	1	0	0	0	0	1	0	0	0	1
V4	1	2	1	0	0	0	3	3	1	3	0	2	3	3
V5	3	1	2	0	0	1	0	1	2	2	2	1	1	2
V6	1	3	1	0	0	0	3	3	1	3	0	2	3	3
V7	3	1	2	0	0	1	0	1	2	1	2	1	1	2
V8	1	1	0	0	0	0	0	0	0	1	0	0	0	1
V9	3	3	3	3	2	2	1	3	2	3	3	3	0	3
V10	0	1	2	1	0	2	1	1	0	0	0	3	1	0
Mean	1.6	1.6	1.6	0.8	0.5	1.0	1.0	1.6	0.9	1.7	1.0	1.7	1.0	1.8

then it will display the information based on the severity level. As it is observed that "V5" is detected with psychological risks in all three sets. An "MD" level of risk is recognized for both anxiety and depression disorder while an "M" risk level is detected with stress set. The predicted risk disorders are supported and regulated by a well-built interpreted engine. The interpreter engine explains the predicted outcome corresponding to all three disorders. Permuted feature importance method, CEM method, and Counterfactual method are the three different explanatory approaches implemented in this module for a comprehensive interpretation of predicted result class for each mental disorder set. Prediction interpretations for each set are presented in context to the detected outcome of volunteer "V5."

5.3. Anxiety Interpreter

Anxiety interpreter acts as an reasoning interface that facilitates an effective explanation of the prediction outcome on the anxiety dataset which is depicted through the three significant methods discussed above. **Figure 14** shows the anxiety interpreter.

TABLE 8 | Irrelevant queries detection and elimination in Q-Prioritization phase.

Anxiety														
Q2	Q4	Q7	Q9	Q15	Q19	Q20	Q23	Q25	Q28	Q30	Q36	Q40	Q41	Pre-Q-Prioritized
Q9	Q40	Q41	Q23	Q15	Q19	Q28	Q7	Q36	Q25	Q20	Q4	Q30	Q2	Post-Q-Prioritized
Stress														
Q1	Q6	Q8	Q11	Q12	Q14	Q18	Q22	Q27	Q29	Q32	Q33	Q35	Q39	Pre-Q-Prioritized
Q18	Q27	Q39	Q11	Q29	Q6	Q35	Q8	Q14	Q33	Q12	Q22	Q32	Q1	Post-Q-Prioritized
Depression														
Q3	Q5	Q10	Q13	Q16	Q17	Q21	Q24	Q26	Q31	Q34	Q37	Q38	Q42	Pre-Q-Prioritized
Q5	Q13	Q26	Q17	Q21	Q34	Q38	Q3	Q5	Q10	Q24	Q31	Q37	Q42	Post-Q-Prioritized

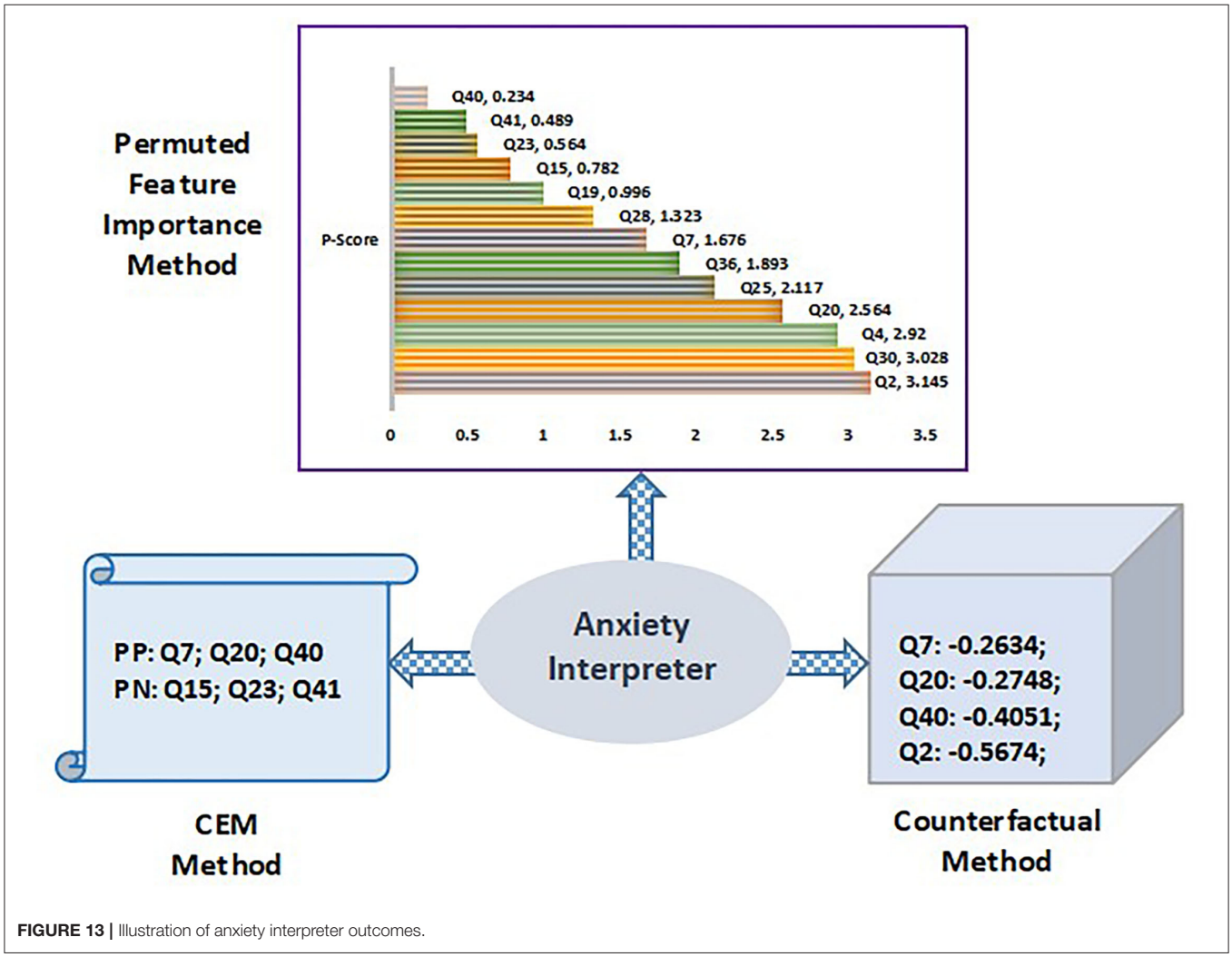


FIGURE 13 | Illustration of anxiety interpreter outcomes.

In context to permuted feature importance method, the P-score value determines the importance of attributes. Q-score value changes with every iteration based on the individual position of attributes. Therefore, *P*-score also updates with every iteration. The value of *P*-score depends on the number of iterations and the accuracy value of all attributes in every round of execution. **Figure 8** shows the ordered ranking of different queries based on their individual the P-score value for the volunteer "V5." It is observed that "Q2" exhibit the highest importance score of 3.145 while "Q40" is the least important attribute with a score of 0.234. It shows the top ranked query attributes that impact the prediction decision making. CEM method is another approach to provide a suitable explanation for the prediction. It denotes the minimal queries to be present for the prediction of a specific class (PP) while simultaneously it also notifies the queries that should be absent during the prediction

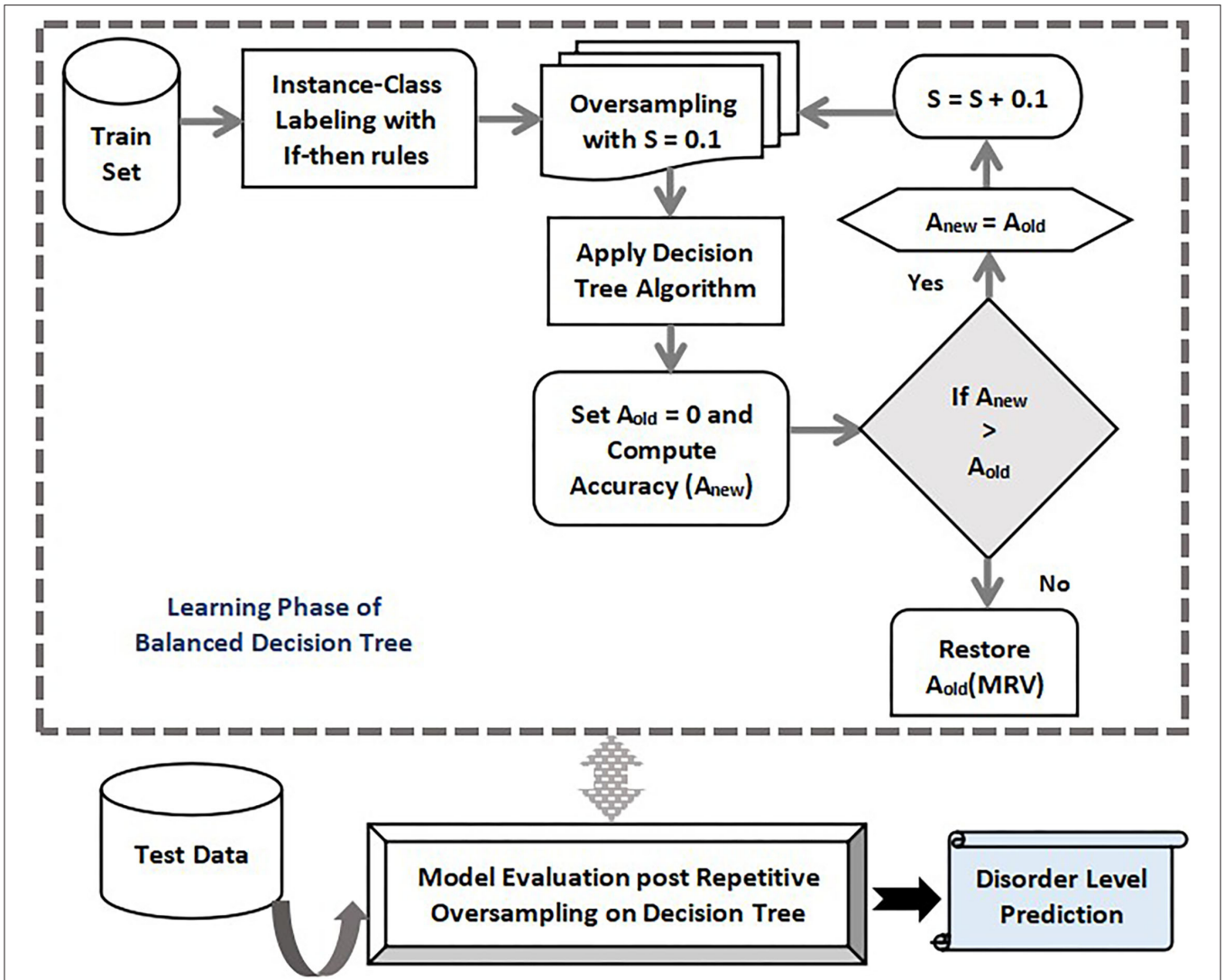


FIGURE 14 | Training and testing phase with proposed balanced decision tree approach.

(PN).The inference of the CEM method is presented in tabular form in **Table 11**. It clearly indicates the scenario where queries Q7, Q20, and Q40 are the PP while Q15, Q23, and Q41 are regarded as PN. Based on the inference, relevant feedback is generated to the user for guidance as shown in **Table 9**.

The counterfactual method is the third interpreter approach used for effective explanation for prediction. As it can be seen that the original prediction for anxiety was "MD." So to downgrade the prediction to "M," the queries Q7, Q20, Q40, and Q2 represent the counterfactuals generated that are listed below. The negative values of the queries denote that the corresponding metric should be reduced. Thus, to reduce the anxiety risk to a mild level, "V5" should control the feeling of panic and constant shakiness. Furthermore, symptoms such as dryness in the mouth and calmness will assist in dealing with anxiety related issues.

Original predicted label: "MD"; Counterfactual prediction: "M"; Q7: –0.2634; Q20: –0.2748; Q40: –0.4051; Q2: –0.5674.

5.4. Stress Interpreter

Stress interpreter forms the second platform to provide an appropriate interpretation of the prediction outcome on the stress dataset. The explanation is shown through the same methods as highlighted above. Stress interpreter and its decision outcomes are highlighted in **Figure 15**.

According to the CEM method, the essential attribute queries (PP) recorded were Q1, Q22, Q29, and Q33 while the non-essential queries needed to be absent (PN) were Q6, Q12, Q14, and Q27. The stress feedback is shown in **Table 10** where the relevant feedback is provided to the user based on the generated "PP" and "PN" values.

The query importance is further validated with the counterfactual method to determine the attributes of the stress set that need to be regulated. The counterfactual results showed that "M" was the actual predicted class level, but it can be further regulated to "N" if certain attributes are properly taken care of. The attributes of impact detected were Q1, Q29,

TABLE 9 | Feedback to anxiety interpreter with CEM method.

Volunteer ID	Queries	PP	PN	Feedback (anxiety)
V5	Q2; Q15; Q23; Q30; Q41	Q7; Q20; Q28; Q40	Q15; Q23; Q41	V5 is detected "Moderate" anxiety risk due to shakiness feeling, scared and panic. However, if he would have experienced faintness, difficulty in swallowing and trembling in hands then the risk would be upgraded to "Severe."

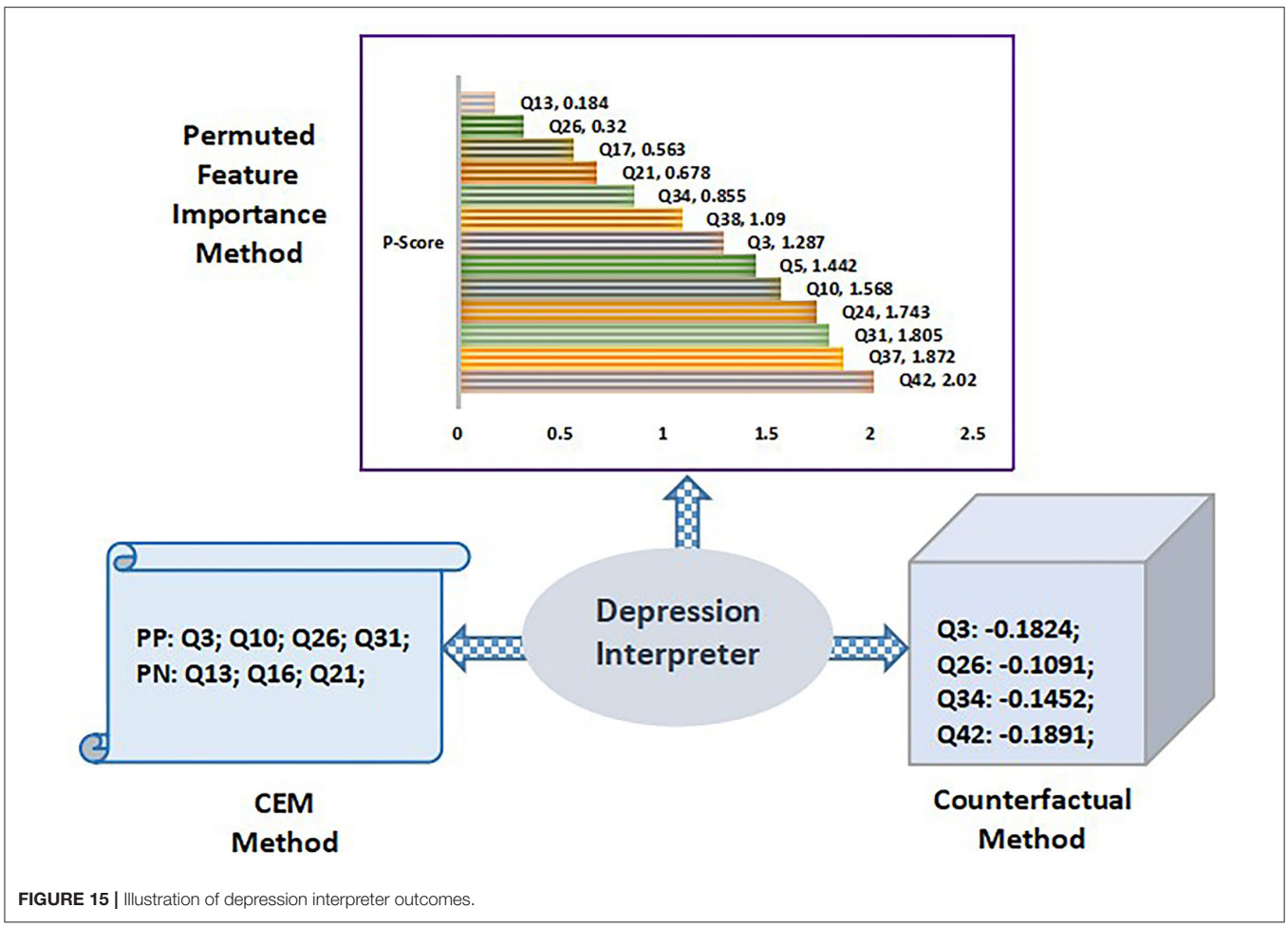


FIGURE 15 | Illustration of depression interpreter outcomes.

TABLE 10 | Feedback to stress interpreter with CEM method.

Volunteer ID	Queries	PP	PN	Feedback (stress)
V5	Q1; Q6; Q11; Q12; Q22 Q27; Q32; Q33; Q39	Q8; Q1; Q14; Q22; Q29; Q35; Q33	Q6, Q12; Q14; Q27	V5 is detected with "Mild" stress risk due to being easily upset, unable to rest and maintain calm with excess tension. Moreover, if he tends to over-react more, get nervous and impatient with frequent irritation then he would develop "Moderate" risk.

TABLE 11 | Feedback to depression interpreter with CEM method.

Volunteer ID	Queries	PP	PN	Feedback (stress)
V5	Q3; Q10; Q16; Q17; Q24 Q26; Q34; Q37; Q42	Q13; Q3; Q21; Q12; Q31; Q26; Q38; Q31	Q13; Q16; Q31	V5 is detected with "Moderate" depression risk due to high negative feeling, lack of objective in life, down-hearted and worthless mindset. However, if he continues to remain sad continuously with no zeal and worth in life feeling then the risk may be uplifted to "Severe."

Q32, and Q33 as listed in **Figure 2**. So as per the counterfactual outcome, volunteer "V5" should avoid being frequently upset for silly reasons and should maintain calm. Furthermore, the usage of sensors and smart gadgets will enable remote monitoring of patients. Original predicted label: "M" Counterfactual prediction: "N" Q1: –0.2284; Q29: –0.1941; Q32: –0.3382; Q33: –0.1674.

5.5. Depression Interpreter

The user interface used in our work to provide an effective explanation for depression prediction is referred to as depression interpreter.

It also utilizes the same three approaches to provide a suitable explanation for prediction. Through permuted feature importance method, the priority ranking of all queries is obtained by their P-score value. The highest P-score of 2.02 is noted for query "Q42" while the least score of 0.184 is recorded for "Q13." So it can be said that "Q42" is the most relevant query that had an impact on the prediction of depression in volunteer "V5." The critical queries required for accurate prediction and queries that should be absent are computed using the CEM approach. It is noted that Q3, Q10, Q26, and Q31 are absolutely required for prediction of the same class level "MD" while queries "Q13," "Q16," and "Q21" should not impact the prediction process. On basis of the CEM method outcome, the feedback table is auto-generated as shown in **Table 11**.

The query importance in depression risk prediction is evaluated with the counterfactual method to determine the queries that require attention. The counterfactual results showed that "MD" was the original predicted depression level which may be upgraded to "M" if four attributes are effectively controlled. These detected attributes were Q3, Q26, Q34, and Q42 as listed in **Table 5**. Hence, a volunteer "V5" should be more positive rather than being down-hearted. A volunteer should be kept busy to avoid anxiety related issues.

Q3: –0.1824; Q26: –0.1091; Q34: –0.1452; Q42: –0.1891;

The performance of the proposed model can be further enhanced in terms of speed and accuracy by taking image based datasets and using deep neural networks. Also, the entire system can be made smart by integrating sensors and actuators so that remote monitoring of patients can be possible. In the future, the model can be further upgraded to include more detailed questionnaires that capture the emotions of patients more accurately. Also, more mental risks types can be embedded in the research work to make the model more generic and broad.

6. CONCLUSION

Psychological health plays a crucial role in the well-being of a person. Any disturbance related to mental disorders may pose serious concerns in the future. Anxiety, stress, and depression contribute to the majority of psychological risks. Thus, it is important to recognize these health concerns and address them at the earliest. Many predictive learning based models are already being developed to predict these mental disorders. But most of these models simply predict the occurrence of such risks without providing any concrete explanation and reasoning. So reliability on these black-box models is questionable. So a more transparent and robust approach should be deployed such that the predictions of a model can be accurately visualized and interpreted. Thus, in our research, an explainable intelligence enabled predictive model is developed to not only accurately detect multiple severity levels of psychological health risks in an individual but also generate a transparent explanation of the contributing factors of a certain prediction. The model used a standard online questionnaire dataset to accumulate responses from online users. The data samples were filtered using a new method called Q-Prioritization through which the least priority queries in the dataset were dropped out. The query optimized dataset was trained with a new repetitive oversampled variant of the decision tree called the balanced decision tree method. After the successful learning phase, it was able to predict psychological issues such as anxiety, stress, and depression on a new test sample. Furthermore, a reasoning engine is used to visualize and interpret the prediction made by the model so as to determine the main contributing symptoms and less essential factors for a certain prediction of mental risks severity level. Permuted feature importance, contrastive explanation, and counterfactuals methods are utilized as constituents of the reasoning engine to empower the model with explainable intelligence. The deployed model gave very impressive results. The mean accuracy of prediction was 98.25% while the mean precision, recall, and F-score metric noted were 0.98, 0.977, and 0.979, respectively. It was observed that without applying Q-Prioritization, the mean accuracy significantly got reduced to 90.25%. The MSE generated with our model was a least of 0.026. Through the reasoning engine, the prediction outcome was more clear and transparent. Risk factors primarily responsible for a certain severity disorder level detection were quite evident and relatable. Based on the outcome, medical experts can suggest suitable recommendations to patients. Thus, the proposed multi-level psychological disorder predictive model driven by explainable intelligence can benefit medical professionals in the accurate diagnosis of patients concerned with psychological health risks. In the future, the model can be further upgraded to include more detailed questionnaires that capture the emotions of patients more accurately. Also, more mental risk types can be embedded in the research work to make the model more generic and broad.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SM, HT, HK, DG, KK, and SP: conceptualization and data collection, methodology, and manuscript drafting. KK: manuscript proofreading. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the research work of Symbiosis Institute of Technology.

REFERENCES

1. Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. *IEEE Trans Affect Comput*. (2017) 10:374–93. doi: 10.1109/TAFFC.2017.2714671
2. Anjume S, Amandeep K, Aijaz A, Kulsum F. Performance analysis of machine learning techniques to predict mental health disorders in children. *Int J Innovat Res Comput Commun Eng*. (2017) 5.
3. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, et al. CP-BDHCA: blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform*. (2021) doi: 10.1109/JBHI.2021.3097237
4. Chatterjee M, Stratou G, Scherer S, Morency LP. Context-based signal descriptors of heart-rate variability for anxiety assessment. In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence: IEEE (2014). p. 3631–5.
5. Vuppalapati C, Raghu N, Veluru P, Khursheed S, A system to detect mental stress using machine learning and mobile development. In: *2018 International Conference on Machine Learning and Cybernetics (ICMLC). Vol. 1*. Chengdu: IEEE (2018). p. 161–6.
6. Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, et al. Depression detection via harvesting social media: a multimodal dictionary learning solution. In: *IJCAI*. Austria. (2017). p. 3838–44.
7. Dabek F, Caban JJ. A neural network based model for predicting psychological conditions. In: *International conference on brain informatics and health*. London: Springer (2015). p. 252–61.
8. Pandya S, Ghayvat H. Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence. *Adv Eng Inform*. (2021) 47:101238. doi: 10.1016/j.aei.2020.101238
9. Pandya S, Patel W. An adaptive approach towards designing a smart health-care real-time monitoring system based on iot and data mining. In: *3rd IEEE International Conference on Sensing technology and Machine Intelligence (ICST-2016)*. (2016).
10. Ghayvat H, Awais M, Gope P, Pandya S, Majumdar S. Recognizing suspect and predicting the spread of contagion based on mobile phone location data (counteract): a system of identifying covid-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence. *Sustain Cities Soc*. (2021) 69:102798. doi: 10.1016/j.scs.2021.102798
11. Pandya S, Patel W, Ghayvat H. NXTGeUH: ubiquitous healthcare system for vital signs monitoring & falls detection. In: *IEEE International Conference, Symbiosis International University*. Pune: IEEE (2018).
12. Thakkar HK, Sahoo PK. Towards automatic and fast annotation of seismocardiogram signals using machine learning. *IEEE Sens J*. (2019) 20:2578–89. doi: 10.1109/JSEN.2019.2951068
13. Thakkar HK, Liao Ww, Wu Cy, Hsieh YW, Lee TH. Predicting clinically significant motor function improvement after contemporary task-oriented interventions using machine learning approaches. *J Neuroeng Rehabil*. (2020) 17:1–10. doi: 10.1186/s12984-020-00758-3
14. Mishra S, Thakkar H, Mallick PK, Tiwari P, Alamri A. A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection. *Sustain Cities Soc*. (2021) 103079. doi: 10.1016/j.scs.2021.103079
15. Thakkar HK, Shukla H, Patil S. A comparative analysis of machine learning classifiers for robust heart disease prediction. *In: 2020 IEEE 17th India Council International Conference (INDICON)*. New Delhi: IEEE (2020). p. 1–6.
16. Dooshima MP, Chidozie EN, Ademola BJ, Sekoni OO, Adebayo IP. A predictive model for the risk of mental illness in Nigeria using data mining. *Int J Immunol*. (2018) 6:5–16. doi: 10.11648/j.iji.20180601.12
17. Srividya M, Mohanavalli S, Bhalaji N. Behavioral modeling for mental health using machine learning algorithms. *J Med Syst*. (2018) 42:1–12. doi: 10.1007/s10916-018-0934-5
18. Alonso SG, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno DC, Nozaleda LM, et al. Data mining algorithms and techniques in mental health: a systematic review. *J Med Syst*. (2018) 42:1–15. doi: 10.1007/s10916-018-1018-2
19. Patel W, Pandya S, Mistry V. i-MsRTRM: developing an IoT based intelligent medicare system for real-time remote health monitoring. In: *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*. Tehri: IEEE (2016). p. 641–5.
20. Sau A, Bhakta I. Screening of anxiety and depression among seafarers using machine learning technology. *Inform Med Unlocked*. (2019) 16:100228. doi: 10.1016/j.imu.2019.100228
21. Saha B, Nguyen T, Phung D, Venkatesh S. A framework for classifying online mental health-related communities with an interest in depression. *IEEE J Biomed Health Inform*. (2016) 20:1008–15. doi: 10.1109/JBHI.2016.2543741
22. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. *Sci Rep*. (2017) 7:1–11. doi: 10.1038/s41598-017-12961-9
23. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Mental Health*. (2016) 3:e21. doi: 10.2196/mental.4822
24. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, et al. Extracting psychiatric stressors for suicide from social media using deep learning. *BMC Med Inform Decis Mak*. (2018) 18:77–87. doi: 10.1186/s12911-018-0632-8
25. Al Hanai T, Ghassemi MM, Glass JR. Detecting depression with audio/text sequence modeling of interviews. In: *Interspeech*. Hyderabad. (2018). p. 1716–20.
26. Ramiandrisoa F, Mothe J, Benamara F, Moriceau V. IRIT at e-Risk 2018. In: *9th Conference and Labs of the Evaluation Forum, Living Labs (CLEF 2018)*. (2018). p. 1–12.
27. Hou Y, Xu J, Huang Y, Ma X. A big data application to predict depression in the university based on the reading habits. In: *2016 3rd International Conference on Systems and Informatics (ICSAI)*. Shanghai: IEEE (2016). p. 1085–9.
28. Leightley D, Williamson V, Darby J, Fear NT. Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort. *J Mental Health*. (2019) 28:34–41. doi: 10.1080/09638237.2018.1521946
29. Young C, Harati S, Ball T, Williams L. S35. Using machine learning to characterize circuit-based subtypes in mood and anxiety disorders. *Biol Psychiatry*. (2019) 85:S310. doi: 10.1016/j.biopsych.2019.03.786
30. Mary SA, Jabasheela L. An evaluation of classification techniques for depression, anxiety and stress assessment. In: *International Conference for Phoenixes on Emerging Current Trends in Engineering and Management (PECTEAM 2018)*. Atlantis: Atlantis Press (2018). p. 64–9.
31. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry*. (2016) 21:1366–71. doi: 10.1038/mp.2015.198
32. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. *Front Psychiatry*. (2017) 8:72. doi: 10.3389/fpsyt.2017.00072
33. Peng Z, Hu Q, Dang J. Multi-kernel SVM based depression recognition using social media data. *Int J Mach Learn Cybern*. (2019) 10:43–57. doi: 10.1007/s13042-017-0697-1
34. Priya A, Garg S, Tigga NP. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Comput Sci*. (2020) 167:1258–67. doi: 10.1016/j.procs.2020.03.442
35. Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S, Tiffin PA. Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. *J Affect Disord*. (2019) 246:857–60. doi: 10.1016/j.jad.2018.12.095
36. Bauer R, Glenn T, Strejilevich S, Conell J, Alda M, Ardau R, et al. Internet use by older adults with bipolar disorder: international survey results. *Int J Bipolar Disord*. (2018) 6:1–7. doi: 10.1186/s40345-018-0127-7
37. Dhaka P, Johari R. Big data application: study and archival of mental health data, using MongoDB. In: *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. Chennai: IEEE (2016) p. 3228–32.
38. Kumar M, Bala A. Analyzing Twitter sentiments through big data. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. New Delhi: IEEE; (2016). p. 2628–31.
39. Furnham A. A Big Five facet analysis of sub-clinical dependent personality disorder (Dutifulness). *Psychiatry Res*. (2018) 270:622–6. doi: 10.1016/j.psychres.2018.10.052

40. Bleidorn W, Hopwood CJ. Using machine learning to advance personality assessment and theory. *Pers Soc Psychol Rev*. (2019) 23:190–203. doi: 10.1177/1088868318772990
41. Sarraf S, Tofighi G. Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In: *2016 Future Technologies Conference (FTC)*. San Francisco, CA: IEEE (2016). p. 816–20.
42. Fiscon G, Weitschek E, Felici G, Bertolazzi P, De Salvo S, Bramanti P, et al. Alzheimer's disease patients classification through EEG signals processing. In: *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. Orlando, FL: IEEE (2014). p. 105–12.
43. Omurca S I, Ekinci E. An alternative evaluation of post traumatic stress disorder with machine learning methods. In: *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*. Madrid: IEEE (2015). p. 1–7.
44. Katsis CD, Katertsidis NS, Fotiadis DI. An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomed Signal Process Control*. (2011) 6:261–8. doi: 10.1016/j.bspc.2010.12.001
45. Saxe GN, Ma S, Ren J, Aliferis C. Machine learning methods to predict child posttraumatic stress: a proof of concept study. *BMC Psychiatry*. (2017) 17:1–13. doi: 10.1186/s12888-017-1384-1
46. Karstoft KI, Galatzer-Levy IR, Statnikov A, Li Z, Shalev AY. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry*. (2015) 15:1–7. doi: 10.1186/s12888-015-0399-8
47. Ray C, Kumar Tripathy H, Mishra S. Assessment of autistic disorder using machine learning approach. In: *International Conference on Intelligent Computing and Communication*. Springer (2019). p. 209–19.
48. Tripathy HK, Mallick PK, Mishra S. Application and evaluation of classification model to detect autistic spectrum disorders in children. *Int J Comput Appl Technol*. (2021) 65:368–77. doi: 10.1504/IJCAT.2021.117286
49. Ray C, Tripathy HK, Mishra S. A review on facial expression based behavioral analysis using computational technique for autistic disorder patients. In: *International Conference on Advances in Computing and Data Sciences*. Springer (2019). p. 450–64.
50. Mishra S, Mohanty S. Integration of machine learning and IoT for assisting medical experts in brain tumor diagnosis. In: *Smart Healthcare Analytics: State of the Art*. Singapore: Springer (2022). p. 133–64.
51. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell*. (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
52. Gadekallu TR, Khare N. Cuckoo search optimized reduction and fuzzy logic classifier for heart disease and diabetes prediction. *Int J Fuzzy Syst Appl*. (2017) 6:25–42. doi: 10.4018/IJFSA.2017040102

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mishra, Tripathy, Kumar Thakkar, Garg, Kotecha and Pandya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 22 December 2021
doi: 10.3389/fpubh.2021.818439

Development and Internal Validation of a Nomogram to Predict Mortality During the ICU Stay of Thoracic Fracture Patients Without Neurological Compromise: An Analysis of the MIMIC-III Clinical Database

Haosheng Wang[1†], *Yangyang Ou*[1†], *Tingting Fan*[2], *Jianwu Zhao*[1], *Mingyang Kang*[1], *Rongpeng Dong*[1] *and Yang Qu*[1*]

[1] *Department of Orthopedics, Second Hospital of Jilin University, Changchun, China,* [2] *Department of Endocrinology, Baoji City Hospital of Traditional Chinese Medicine, Baoji, China*

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Jincheng Wang, Nanjing Medical University, China
Praveen Kumar, VIT University, India

***Correspondence:**
Yang Qu
quy@jlu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 November 2021*
Accepted: *03 December 2021*
Published: *22 December 2021*

Citation:
Wang H, Ou Y, Fan T, Zhao J, Kang M, Dong R and Qu Y (2021) Development and Internal Validation of a Nomogram to Predict Mortality During the ICU Stay of Thoracic Fracture Patients Without Neurological Compromise: An Analysis of the MIMIC-III Clinical Database. Front. Public Health 9:818439. doi: 10.3389/fpubh.2021.818439

Background: This study aimed to develop and validate a nomogram for predicting mortality in patients with thoracic fractures without neurological compromise and hospitalized in the intensive care unit.

Methods: A total of 298 patients from the Medical Information Mart for Intensive Care III (MIMIC-III) database were included in the study, and 35 clinical indicators were collected within 24 h of patient admission. Risk factors were identified using the least absolute shrinkage and selection operator (LASSO) regression. A multivariate logistic regression model was established, and a nomogram was constructed. Internal validation was performed by the 1,000 bootstrap samples; a receiver operating curve (ROC) was plotted, and the area under the curve (AUC), sensitivity, and specificity were calculated. In addition, the calibration of our model was evaluated by the calibration curve and Hosmer-Lemeshow goodness-of-fit test (HL test). A decision curve analysis (DCA) was performed, and the nomogram was compared with scoring systems commonly used during clinical practice to assess the net clinical benefit.

Results: Indicators included in the nomogram were age, OASIS score, SAPS II score, respiratory rate, partial thromboplastin time (PTT), cardiac arrhythmias, and fluid-electrolyte disorders. The results showed that our model yielded satisfied diagnostic performance with an AUC value of 0.902 and 0.883 using the training set and on internal validation. The calibration curve and the Hosmer-Lemeshow goodness-of-fit (HL). The HL tests exhibited satisfactory concordance between predicted and actual outcomes ($P = 0.648$). The DCA showed a superior net clinical benefit of our model over previously reported scoring systems.

Conclusion: In summary, we explored the incidence of mortality during the ICU stay of thoracic fracture patients without neurological compromise and developed a prediction

model that facilitates clinical decision making. However, external validation will be needed in the future.

Keywords: intensive care units, nomogram, spinal fractures, prediction model, mortality

BACKGROUND

A spinal fracture is a dislocation or fracture of the vertebrae with an annual incidence of 26 per 100 000 (1), accounting for ∼23.3% of all trauma patients (2). The most prevalent spinal fractures occur at the level of the lumber spine, followed by the thoracic spine (3). Motor vehicle accidents (MVC) and falls from heights are the most common causes of spinal fracture; the annual incidence of spinal fracture has increased, with the reported increase in motor vehicle crashes (4). In China, the incidence rate of spinal fractures in 2007 was twice that in 2001 (3).

Numerous studies have investigated the long-term prognosis of spinal fractures using data from large patient populations in recent years. In patients with thoracolumbar fractures exhibiting mild symptoms, the visual analog scale (VAS) scores and the Roland Morris Disability Questionnaire demonstrated poor results and a poor outcome in 6% of patients 10 years after non-operative treatment (5). Furthermore, for patients exhibiting more severe symptoms requiring surgery, only 50% reportedly return to their original jobs postoperatively (6).

Meanwhile, long-term mortality associated with spinal fractures is significantly increased in post-traumatic patients compared to those with no history of trauma (7), especially in osteoporotic patients (8, 9). It has also been shown that old age and male gender increase the long-term mortality risk after thoracolumbar fracture (10). Interestingly, studies have shown that trauma patients in the ICU have high mortality rates, as high as 31% (11). Notwithstanding that substantial progress has been made in trauma advanced life support over the years, the mortality rates are still high (12). Patients with concomitant spinal cord injury are more likely than those without neurological damage to present with multiorgan injury and die from secondary infection (13). Most importantly, many patients experience death prior to hospitalization due to severe spinal cord injuries (14). To the best of our knowledge, the prognosis of acute severe spine fracture patients in ICU has been largely unexplored, and current research hot spots still focus on osteoporosis, bone cement, biomechanical analysis, and so on. Meanwhile, the interest of intensivists in trauma patients has mostly focused on rib fractures and pelvic fractures (15–17). Due to the lack of adequate attention to such patients, a large number of high-risk patients are not identified in time for admission, which often leads to poor prognosis and even higher mortality. Little is currently known about the risk factors of poor prognosis in ICU patients with thoracic spine fractures, which has resulted in high mortality rates. Accordingly, we selected this specific patient population for our study.

It is essential to develop a prediction model based on routine clinical and laboratory parameters to ensure that it can be easily implemented during clinical practice. It has been established that nomograms can provide evidence-based and personalized risk estimates and contribute to clinical management and prognosis evaluation (18–21). This study aimed to develop a prediction model to predict mortality during the ICU stay of thoracic fracture patients without neurological compromise based on the Medical Information Mart for Intensive Care III (MIMIC-III) (22) clinical database.

MATERIALS AND METHODS

Data Source

Our data were derived from the MIMIC-III database, a database established and open source by the Massachusetts Institute of Technology (MA, USA) containing information on more than 58,000 patients that attended the Beth Israel Deaconess Medical Center. We completed a web course offered by the National Institutes of Health (NIH) and were granted access to the MIMIC-III database (certification number: 42442549). The data were extracted from the MIMIC-III database using structure query language (SQL) with pgAdmin4 PostgreSQL 9.6.

Study Population

Structured Query Language (SQL) with PostgreSQL (version 9.6 University of California, Berkeley) was used to extract information associated with each patient's unique HADM_ID from the MIMIC-III database. Through International Classification of Diseases 9th Edition (ICD-9) code = 8,052, we obtaining 381 patient. For patients with multiple admissions, we retained information only on the patient's first admission to the ICU. Patients with missing data >20% for laboratory tests ($n = 5$) and ital signs ($n = 1$) were excluded. Finally, 298 patients were included in the study (**Figure 1**).

Clinical Variables and Definition

Using the patient's HADM_ID and ICUSTAY_ID as described above, we extracted the following data: demographics, vital signs, laboratory tests, comorbidities, and scoring systems. Among these, demographics included age and gender; vital signs included heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean blood pressure (MBP), respiratory rate (RR), temperature, SpO_2, bicarbonate; laboratory tests included blood urea nitrogen (BUN), chloride, creatinine, glucose, hemoglobin, potassium, platelet, partial thromboplastin time (PTT), prothrombin time (PT), sodium, white blood cell (WBC), red blood cell (RBC); comorbidities included congestive heart failure, cardiac arrhythmias, liver disease, coagulopathy, fluid-electrolyte disorders, hypertension, renal failure, obesity, chronic pulmonary; scoring systems include scoring systems included simplified acute physiology score II

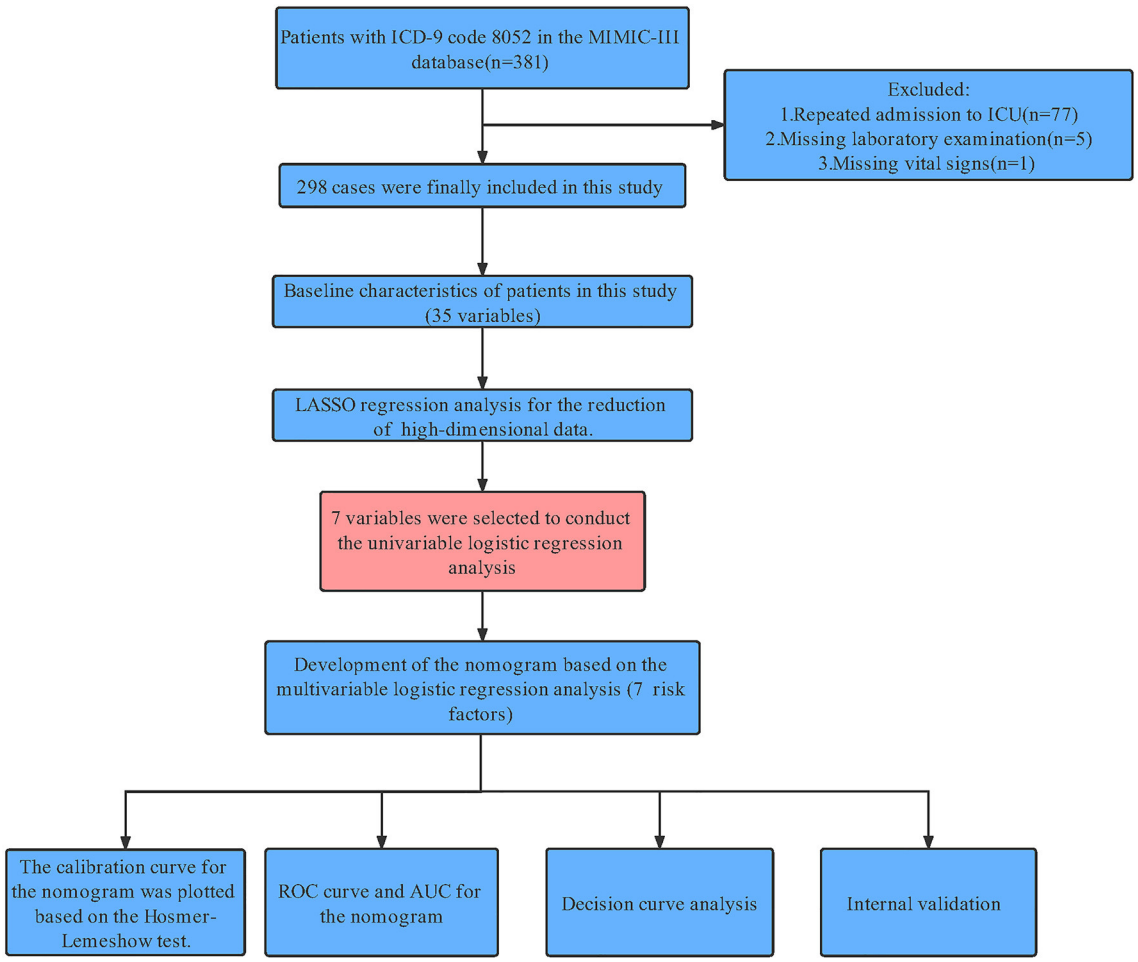


FIGURE 1 | Workflow of the study. MIMIC-III, Medical Information Mart for Intensive Care III; ICU, intensive care unit; LASSO, least absolute shrinkage and selection operator; AUC, area under the curve; HL test, Hosmer-Lemeshow goodness-of-fit test; DCA, decision curve analysis.

(SAPS II), sequential organ failure as severity (sofa) score, oxford acute severity of illness score (OASIS), and the Glasgow coma scale (GCS). Indicators with >20% missing data were removed, such as height, weight, calcium; A deletion was also made for some complications that were not present in all patients, such as peptic ulcer, lymphoma, metastatic cancer. All the variables mentioned above were collected within 24 h of patients' admission.

Statistical Analysis

We used the median and interquartile range to express continuous variables, while Wilcoxon's rank-sum test was selected for comparison between two groups; The categorical variables were expressed as frequency and proportion. Chi-square tests or Fisher's exact test was chosen for inter-group comparison according to the situation. We selected the bootstrap method for internal validation based on the number of patients, in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines (23). We used the neighbor interpolation method in the MICE R package (24) to fill in missing data. Then, the least absolute shrinkage and selection operator (LASSO) expression was used for screening predictors of mortality (25). For the cross-validation results, we selected lambda = min to determine the final candidate variables (26, 27). The multivariate logistic regression model was established using these variables, and multicollinearity was evaluated by variable inflation factors (VIF). The area under the receiver operating curve (AUC), sensitivity and specificity were used to evaluate the model's performance. The Youden index determined the best cutoff point. Finally, the nomogram was plotted using the R package "regplot." The calibration C index (bootstrap resampling 1,000 times) (28), the calibration curve (relationship between observation probability and prediction probability), Hosmer-Lemeshow goodness of fit test (HL test), and brier score were used to evaluate the degree of consistency between observed and predicted outcomes. Decision curve analysis (DCA) was used to assess the net clinical benefit (29). All statistical analyses were completed using R language (version 3.6.3); a $p < 0.05$ was statistically significant.

TABLE 1 | Baseline characteristics of in-hospital death and survival groups.

Variable	Total (n = 298)	Survival (n = 270)	Death (n = 28)	P-value
Age, years	53.50 [35.00, 72.00]	50.00 [32.25, 68.00]	76.50 [66.25, 80.00]	<0.001
Female, *n* (%)	103 (34.6)	89 (33.0)	14 (50.0)	0.111
HR, beats/min	146.50 [74.25, 219.75]	145.50 [75.25, 220.75]	149.00 [71.50, 200.75]	0.744
SBP, mmHg	120.22 [111.13, 133.15]	121.19 [111.81, 134.21]	109.43 [102.06, 118.81]	<0.001
DBP, mmHg	63.33 [57.08, 70.51]	64.40 [57.75, 70.83]	58.86 [53.31, 63.82]	0.019
MBP, mmHg	80.25 [73.24, 87.63]	81.04 [74.10, 88.55]	74.01 [70.58, 77.88]	0.001
RR, breaths/min	17.84 [15.88, 20.95]	17.65 [15.79, 20.63]	20.75 [17.42, 23.28]	0.001
Temperature, °C	37.00 [36.56, 37.41]	37.05 [36.60, 37.43]	36.66 [36.16, 37.36]	0.041
PO2, mmHg	97.85 [96.41, 99.05]	97.87 [96.41, 99.07]	97.54 [96.65, 98.99]	0.714
Bicarbonate, mmol/L	23.00 [21.00, 26.00]	24.00 [21.00, 26.00]	21.00 [17.75, 24.00]	0.007
BUN, mg/dl	15.00 [12.00, 20.00]	15.00 [12.00, 20.00]	16.00 [14.00, 24.25]	0.174
Chloride, mEq/L	107.00 [103.00, 109.75]	106.00 [103.00, 109.00]	108.50 [101.00, 112.75]	0.129
Creatinine, mg/dL	0.90 [0.70, 1.10]	0.90 [0.70, 1.10]	0.90 [0.70, 1.20]	0.955
Blood glucose, mEq/L	136.00 [113.00, 161.00]	134.00 [112.00, 160.75]	147.50 [133.00, 172.50]	0.066
Hemoglobin, g/dl	12.35 [11.00, 13.78]	12.40 [11.12, 13.80]	10.90 [9.17, 13.33]	0.01
Platelet, K/μL	221.50 [175.00, 282.75]	224.00 [182.00, 288.00]	178.50 [135.75, 239.25]	0.003
Potassium, mEq/L	4.10 [3.80, 4.50]	4.10 [3.73, 4.40]	4.15 [3.80, 4.90]	0.519
PT, s	13.30 [12.50, 14.50]	13.30 [12.50, 14.38]	15.05 [13.18, 18.05]	0.002
PTT, s	26.30 [23.83, 29.80]	26.05 [23.72, 29.20]	33.50 [27.42, 43.73]	<0.001
RBC, m/μL	4.00 [3.59, 4.47]	4.02 [3.63, 4.49]	3.75 [2.92, 4.26]	0.005
Sodium, mEq/L	140.00 [138.00, 142.00]	140.00 [138.00, 141.00]	141.00 [136.00, 142.25]	0.413
WBC, K/μL	14.35 [10.50, 18.30]	14.25 [10.50, 18.25]	15.05 [11.45, 19.15]	0.541
Congestive heart failure, *n* (%)	20 (6.7)	18 (6.7)	2 (7.1)	1
Cardiac arrhythmias, *n* (%)	43 (14.4)	33 (12.2)	10 (35.7)	0.002
Hypertension, *n* (%)	8 (2.7)	7 (2.6)	1 (3.6)	1
Chronic pulmonary, *n* (%)	34 (11.4)	30 (11.1)	4 (14.3)	0.849
Renal failure, *n* (%)	10 (3.4)	9 (3.3)	1 (3.6)	1
Liver disease, *n* (%)	7 (2.3)	6 (2.2)	1 (3.6)	1
Coagulopathy, *n* (%)	14 (4.7)	10 (3.7)	4 (14.3)	0.04
Obesity, *n* (%)	11 (3.7)	9 (3.3)	2 (7.1)	0.623
Fluid-electrolyte disorders, *n* (%)	61 (20.5)	48 (17.8)	13 (46.4)	0.001
OASIS	32.00 [26.00, 38.00]	31.00 [25.00, 37.00]	41.00 [36.75, 47.25]	<0.001
GCS	15.00 [14.00, 15.00]	15.00 [14.00, 15.00]	15.00 [14.00, 15.00]	0.331
SOFA	3.00 [1.00, 5.00]	2.00 [1.00, 4.00]	6.00 [3.75, 8.25]	<0.001
SAPSII	28.00 [19.00, 37.00]	26.50 [18.00, 35.00]	46.00 [37.75, 52.25]	<0.001

HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; MBP, mean blood pressure; RR, respiratory rate; BUN, blood urea nitrogen; PT, prothrombin time; PTT, partial thromboplastin time; RBC, red blood cell; WBC, white blood cell; SAPSII, simplified acute physiology score II; SOFA, sequential organ failure; OASIS, oxford acute severity of illness score; GCS, Glasgow coma scale.

RESULTS

Patient Characteristics

Two hundred and ninety eight patients were finally included in the study, with an average age of 53.5 years. 34.6% (n = 103) were female, with a mortality rate of 13.6% while male patients accounted for 65.4% (n = 195), with a mortality rate of 7.2%. **Table 1** compares the differences in characteristics between the in-hospital death group and the survival group. Compared with the survival group, patients in the death group were older, had lower body temperature and blood pressure, faster respiratory rate, and lower bicarbonate. Meanwhile, lower hemoglobin and platelet counts were observed in the death group, while PT and PTT values were relatively higher, indicating poor coagulation function. The OASIS, SOFA, and SAPS II scores were significantly higher, with a higher prevalence of cardiac arrhythmias, coagulopathy, and fluid-electrolyte disorders.

Characteristics Selection and Development of a Nomogram

Out of 35 variables, seven remained in the lasso logistic regression model based on the binomial deviance minimum criteria (ratio 5:1) (**Figure 2**). The final seven variables included in the multivariate logistic regression, were: age (OR: 1.02; 95% CI 0.99–1.06), OASIS score (OR: 1.08; 95%CI 1.00–1.17), SAPS II score

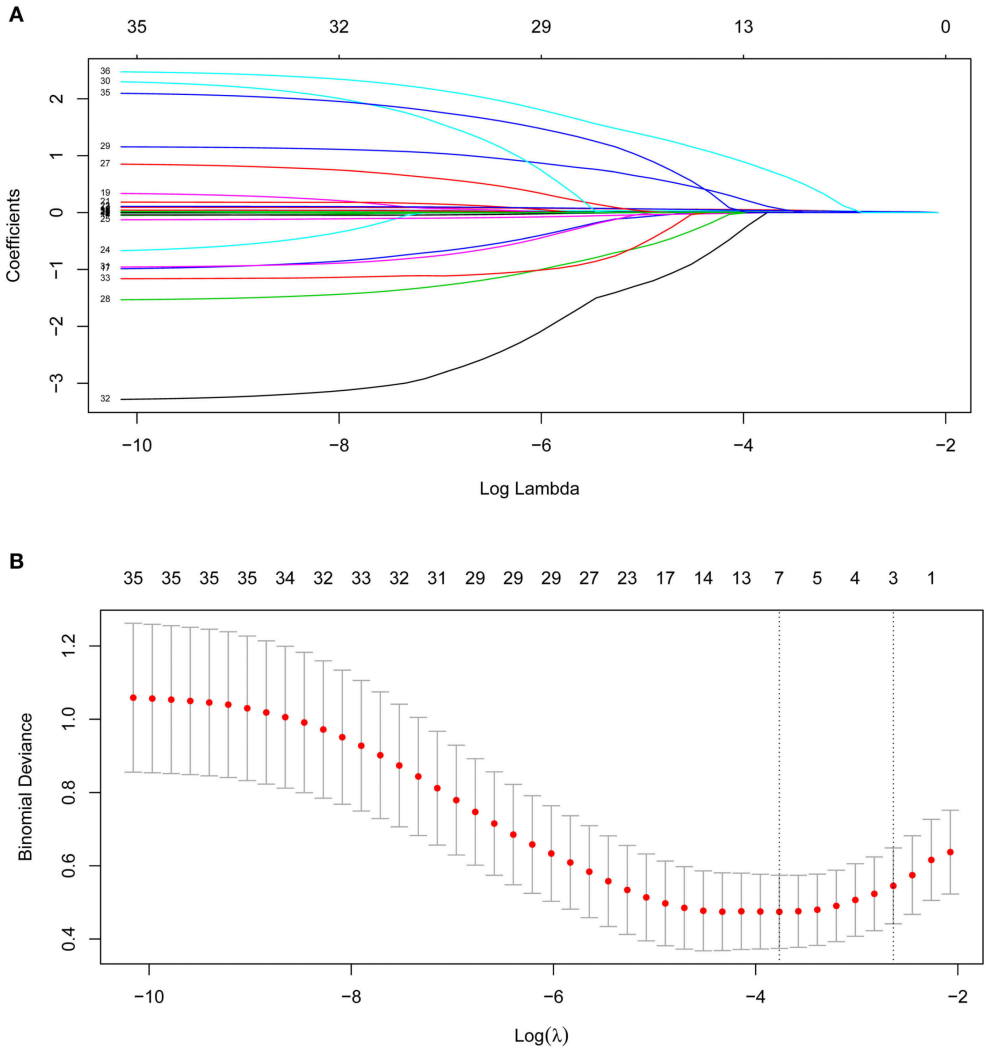


FIGURE 2 | Clinical variables were selected using the lasso logistic regression model. **(A)** Tuning parameter (λ) selection using LASSO penalized logistic regression with 10-fold cross-validation. **(B)** LASSO coefficient profiles of the radiomic features.

(OR: 1.03; 95% CI 0.98–1.08), RR (OR: 1.07; 95% CI 0.94–1.20), PTT (OR: 1.08; 95%CI 1.03–1.13), cardiac arrhythmia (OR: 1.44; 95%CI 0.43–4.62), and fluid-electrolyte disorders (OR: 4.40; 95% CI 1.62–12.54) (**Table 2**). Based on this model, we constructed a nomogram to predict mortality in ICU patients with thoracic spine fractures without neurological injuries (**Figure 3**).

Apparent Performance of the Nomogram and Web Calculator

Our model yielded an AUC value of 0.902 (95% CI 0.849–0.959), with a C-index of 0.883 after 1,000 bootstrap resampling internal validations. According to the Youden index, the optimal cutoff value was 18.45%, with a sensitivity and specificity of 0.870 and 0.786, respectively (**Figure 4**). Meanwhile, the calibration curve showed a good fit during internal validation (**Figure 5**), while the HL test showed that our predicted and observed values were close ($P = 0.648$); The Brier score was 0.0543 and 0.0623 after bootstrap correction. The ROC values of the scoring systems SAPS II and OASIS incorporated into the regression were 0.856 and 0.837, respectively, suggesting that our model exhibited better predictive performance than scoring systems commonly used clinically. To facilitate clinical use, we constructed a web calculator (https://ouyyjlueducn.shinyapps.io/dynnomapp/) based on the model.

Clinical Practice

DCA of the nomogram was performed (**Figure 6**). The blue curve in the figure represents that all the patients received intervention, the straight purple line represents that none patients receive the intervention, and the red curve represents the clinical benefit of our model. For our model, when the predicted probability threshold was set to 0.065%, the net clinical benefit was 9.39%. Our results showed that our model had a superior net clinical benefit than the OASIS and SAPS II scoring systems.

TABLE 2 | Multivariate regression model based on LASSO regression results.

Variables	Multivariable logistics model		
	Coefficients	**OR (95%CI)**	***P*-value**
Age	0.02334	1.02 (0.99–1.06)	0.1627
OASIS	0.074	1.08 (1.00–1.17)	0.069
SAPSII	0.03004	1.03 (0.98–1.08)	0.2284
RR	0.06601	1.07 (0.94–1.20)	0.288
PTT	0.07398	1.08 (1.03–1.13)	0.0028
Cardiac arrhythmias	0.36681	1.44 (0.43–4.62)	0.5394
Fluid-electrolyte disorders	1.48207	4.40 (1.62–12.54)	0.0041

OASIS, oxford acute severity of illness score; SAPSII, simplified acute physiology score II; RR, respiratory rate; PTT, partial thromboplastin time.

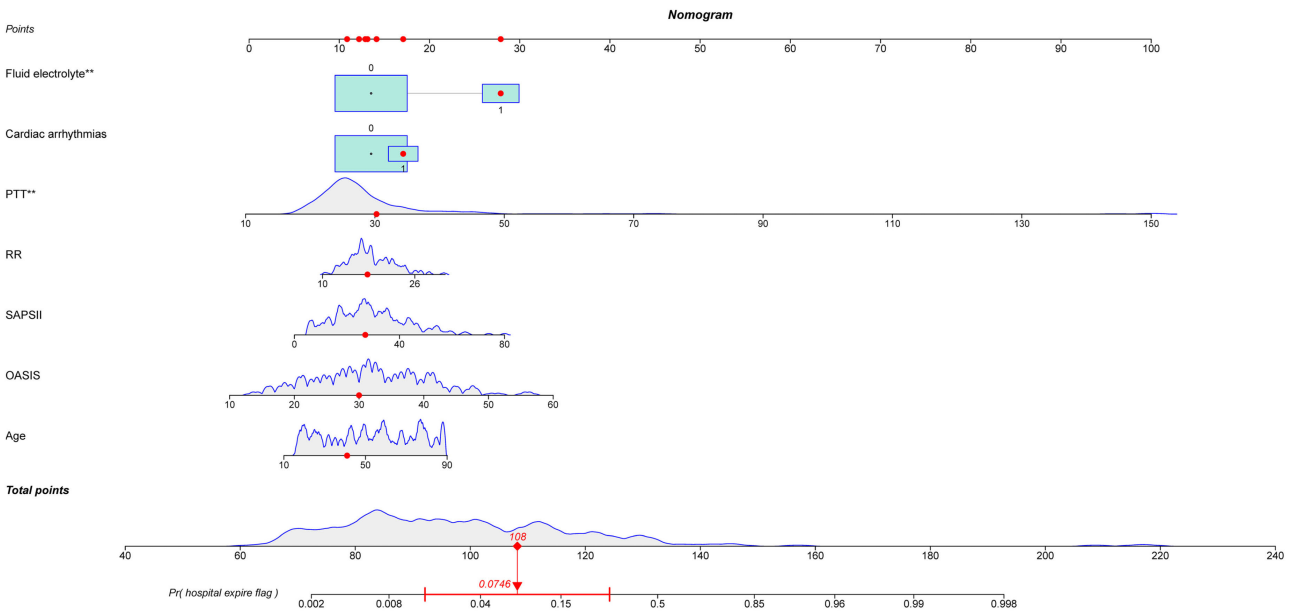


FIGURE 3 | Nomogram to predict the risk of in-hospital mortality in patients with thoracic spine fractures without neurological injury in the ICU. PTT, partial thromboplastin time; RR, respiratory rate; SAPS II, simplified acute physiology score II; OASIS, oxford acute severity of illness score. ** means $p < 0.01$.

DISCUSSION

Fractures have become a major public health problem in most countries (30), especially spine fractures, which account for approximately 46% of spinal injuries in severe trauma (31, 32). In this regard, patients that sustain spine fractures have the poorest functional prognosis and potential to return to work (33). Although many studies have discussed the long-term prognosis of spinal fractures, to the best of our knowledge, this is the first study to investigate the mortality risk of spinal fracture patients without neurological impairment. The injury severity score (ISS) has previously been used to guide the risk stratification of trauma patients; however, some studies have pointed out that this scoring system was not reliable (34) since it did not consider the characteristics of patients with spinal trauma. We summarized the related works in **Table 3**. Previous methods of classifying spinal trauma have been most relevant in guiding surgical treatment (35), and commonly include the Trauma and Injury Severity Score (ISS) and the New Trauma and Injury Severity Score (NISS). However, none of these assessment methods take into account the individual spinal trauma patient, and our work fills a gap in the risk stratification of patients with severe spinal trauma. After internal validation of the model, we found that the predictions of the model were in good agreement with the actual results. More interestingly, the constructed model had better discrimination and net clinical benefit compared to the previously used OASIS and SAPSII systems. Also, based on the reviewers' comments, we have done more work. To facilitate clinical use, we constructed a web calculator (https://ouyyjlueducn.shinyapps.io/dynnomapp/) based on the model. In summary, we provide an easy-to-use model for this group of patients, which can identify high-risk patients early, take appropriate interventions early, and reduce poor prognosis and in-hospital mortality.

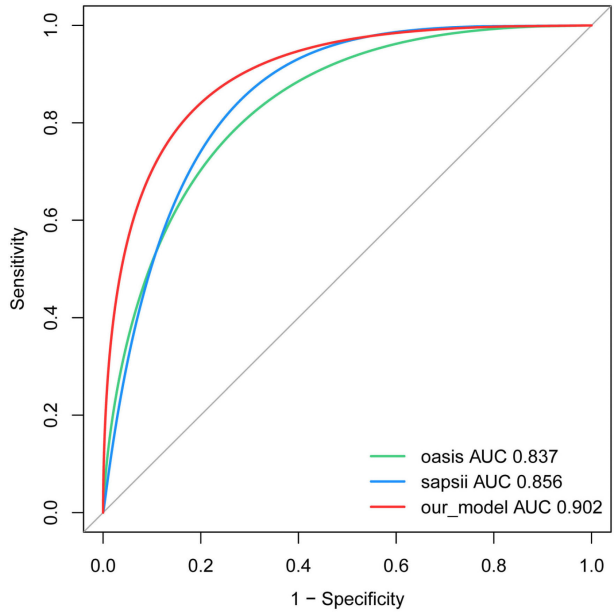


FIGURE 4 | Receiver operating characteristic curve of the nomogram. AUC, area under curve; OASIS, oxford acute severity of illness score; SAPS II, simplified acute physiology score II.

With the rapid development of computer technology in recent years, the combination of machine learning technology and medical practice has become a major trend (19–21, 26, 27). Along with the continuous innovation of convolutional algorithms, from LeNet by Lecun et al. (36) to ResNet by He et al. (37), computer-aided decision making through imaging has become a hot topic in medical research, such as prediction of BMI by facial image features to predict BMI (38) and fundus images to predict diabetic retinopathy (39). In addition, the prevalence of electronic medical records and the establishment of large medical databases have also provided the basis for research on clinical problems, and the combination with machine learning has shown remarkable performance in predicting the occurrence and prognosis of diseases (40, 41). Unfortunately, however, there is still a lack of research in the current field for our patient population. It is therefore necessary to develop an easy-to-use and reliable tool to inform clinical practice. In the present study, we established a nomogram consisting of seven predictors: age, OASIS score, SAPS II score, RR, PTT, cardiac arrhythmias, fluid-electrolyte disorders. Multiple indicators used during the validation process, including AUC, calibration curve, HL test and DCA, indicated that our nomogram yielded satisfactory prediction results.

As seen in **Table 1**, significant differences in coagulation markers PT, PTT, and coagulopathy were found between the death and the survival groups, especially PTT, which was an important predictor of mortality. Trauma-induced coagulopathy (TIC) has been reported to play an important role in trauma healing (42), and an increasing body of evidence suggests that uncontrolled hemorrhage is a preventable cause of death in the early stages after injury (43–45), which has been reported to be as high as 40–80% (46). The incidence rate of TIC often correlates with the severity of tissue injury (47) since tissue injury releases a large number of damage-associated molecular patterns (DAMPs), which promote multiple inflammatory pathways and thereby affect coagulation (42). Endothelial dysfunction has been reported to participate in TIC (48), and plasma samples from severely injured patients on admission exhibited increased levels of syndecan-1, which correlated with increased activated protein C (APC), prolonged PTT, and elevated epinephrine levels (49). Furthermore, platelet defects and dysfunction are also important contributors to TIC (50). In our study, even though the platelet count in the deceased group was still within the normal range, a significant decrease was observed (**Table 2**), suggesting an association between a relatively low platelet count and increased mortality, consistent with previous studies (51, 52). Similarly, a prospective study demonstrated that 91% of patients with severe injuries had platelet dysfunction despite normal platelet counts (53).

The most common arrhythmia in patients with chest trauma or surgery is atrial fibrillation, often associated with longer ICU stay and higher mortality (54). On the one hand, most patients with pre-existing atrial fibrillation receive anticoagulant therapy. When combined with previously described coagulation dysfunction, fatal bleeding can result from small traumas (55). On the other hand, patients with new-onset atrial fibrillation have higher in-hospital mortality than patients with previous atrial fibrillation (56) reportedly. The mortality among patients with arrhythmia in our study group was 23.3% ($n = 10/43$), much higher than that of patients without arrhythmia. Moreover, a significant difference in the number of patients with arrhythmia was found between the survival and death groups ($P < 0.01$).

It is widely acknowledged that the onset of arrhythmias is highly correlated with electrolyte disturbances (54). Many studies have demonstrated that electrolyte disturbances and vertebral fractures are associated with higher mortality, while elderly patients with fractures are more likely to have hyponatremia (57–59). Interestingly, in our study, albeit patients in the deceased group were much older than those in the survival group, no significant difference in sodium levels was found between the two groups. However, our study also confirmed higher mortality in patients who develop fluid-electrolyte disorders, emphasizing the need for early and effective fluid management in this subset of patients (60).

An analysis of nationwide patients with vertebral fractures in Japan revealed that advanced age is a significant risk factor for complications (OR 1.38) (61). We consistently found that advanced age patients were more likely to experience in-hospital death, which may be associated with sarcopenia (62), poor nutritional status (63), and development of fluid-electrolyte disorders, as mentioned earlier. However, we found that gender was not a relevant factor for in-hospital death in our study, which was inconsistent with the literature. This discrepancy could be accounted for by the fact that the subjects of our research were ICU patients, which were critically ill, while gender had more influence in long-term prognosis (10).

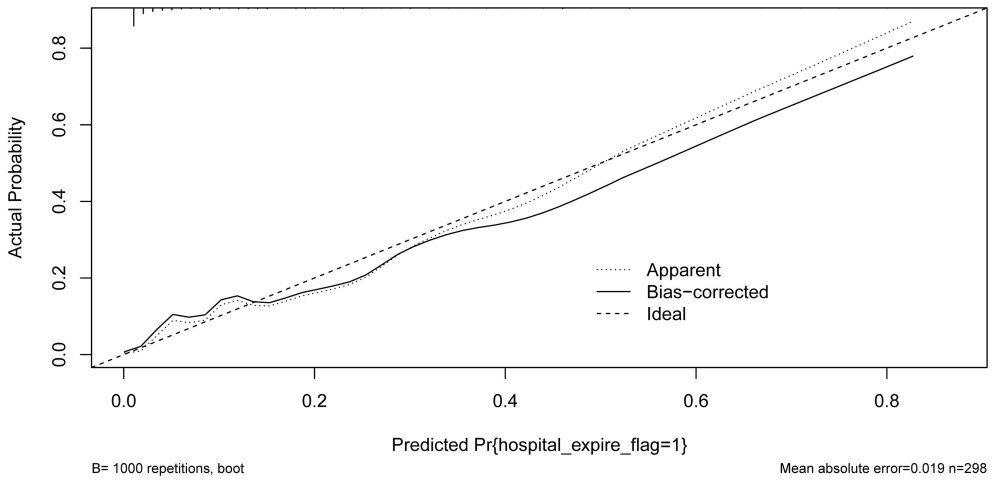


FIGURE 5 | Calibration curves of the predicted nomogram. The dashed line represents the original performance, and the solid dashed line represents the performance during internal validation by bootstrapping (B = 1,000 repetitions). Results of the Hosmer-Lemeshow test demonstrate that the P-value was 0.648.

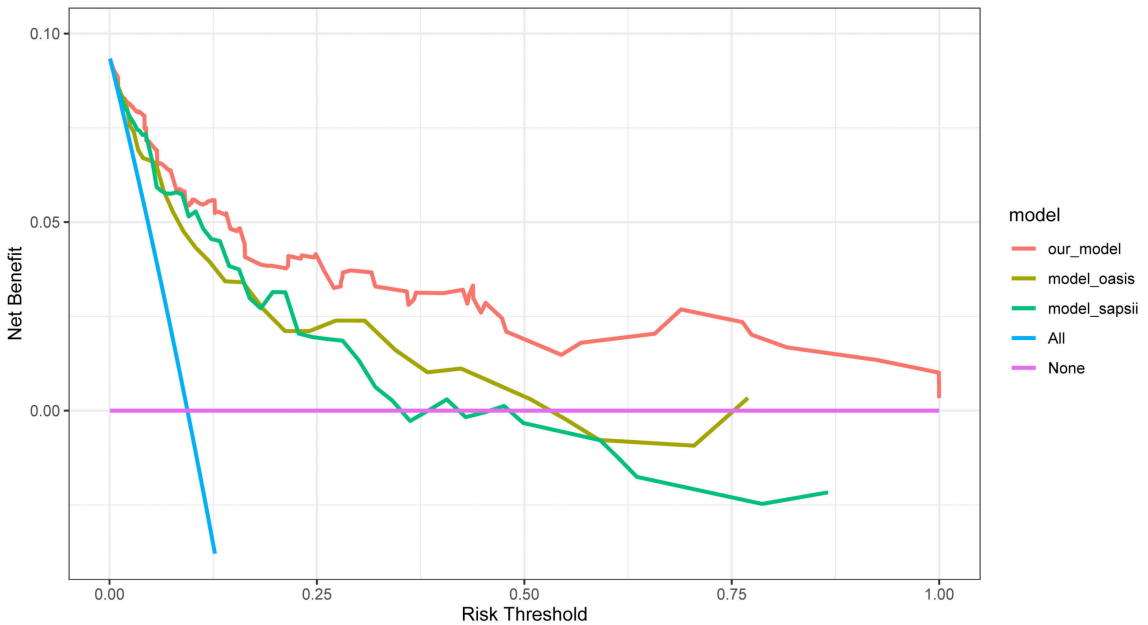


FIGURE 6 | DCA for our model, OASIS and SAPS II. The y-axis measures the net benefit. DCA, decision curve analysis; OASIS, oxford acute severity of illness score; SAPS II, simplified acute physiology score II.

Due to its simplicity and ease of observation, the respiratory rate is one of the indicators traditionally used for the early identification of high-risk patients after trauma (64). A respiratory rate >20 was an important indicator in new evaluation criteria for trauma patients (65). Consistently, we found that tachypneic patients with a respiratory rate within 24 h of admission are more likely to experience in-hospital mortality and require early intervention. Two severity scoring systems, SAPS II and OASIS were also included in the prediction model of this study. At present, much controversy surrounds the predictive accuracy of these two scoring systems in orthopedic trauma patients (66, 67). In our present study, both scoring systems were integrated into our nomogram. The ROC plots showed that our model outperformed these two scoring systems in discrimination (**Figure 3**). At the same time, DCA exhibited greater net clinical benefit than these two scoring systems (**Figure 5**).

There are still limitations that need to be considered. First, these data were from a public database spanning 2007–2014. Therefore, the model needs external validation from different medical Institutions. Second, because missing data is >20% in the dataset, there is a lack of some important clinical parameters. Finally, although nomogram is already widely used

TABLE 3 | The summary of the previous related work.

ID	Title	Author	Journal	Public date
1	Thoracic spine fracture in the panscan era	Remy Bizimungu et al.	Ann Emerg Med	2020 Aug
2	Early risk stratification of in hospital mortality following a ground level fall in geriatric patients with normal physiological parameters	Nasim Ahmed et al.	Am J Emerg Med	2020 Dec
3	Mortality and cause of death in patients with vertebral fractures: a longitudinal follow-up study using a national sample cohort	Hyo Geun Choi et al.	Spine (Phila Pa 1976)	2020 Mar
4	Predicting survival in older patients treated for cervical spine fractures: development of a clinical survival score	Darryl Lau et al.	Spine J	2019 Mar
5	Long-term post-traumatic survival of spinal fracture patients in northern Finland	Ville Niemi-Nikkola et al.	Spine (Phila Pa 1976)	2018 Dec
6	Spinal fractures in older adult patients admitted after low-level falls: 10-year incidence and outcomes	Randeep S. Jawa et al.	J Am Geriatr Soc	2017 May
7	Height loss in older women: risk of hip fracture and mortality independent of vertebral fractures	Teresa A. Hillier et al.	J Bone Miner Res	2012 Jan
8	Mortality and incident vertebral fractures after 3 years of follow-up among geriatric patients	H. C. van der Jagt-Willems et al.	Osteoporos Int	2013 May
9	Predicting in-hospital mortality in elderly patients with cervical spine fractures: a comparison of the Charlson and Elixhauser comorbidity measures	Mariano E. Menendez et al.	Spine (Phila Pa 1976)	2015 Jun
10	Characteristics and outcomes of hospitalized patients with vertebral fragility fractures: a systematic review	Terence Ong et al.	Age Aging	2018 Jan
11	Mortality after vertebral fractures in a Japanese population	Yuzo Ikeda et al.	J Orthop Surg (Hong Kong)	2010 Aug
12	Is radiographic vertebral fracture a risk factor for mortality?	Daniel W. Trone et al.	Am J Epidemiol	2007 Nov
13	Long-term morbidity and mortality after a clinically diagnosed vertebral fracture in the elderly–a 12- and 22-year follow-up of 257 patients	R. Hasserius et al.	Calcif Tissue Int	2005 Apr
14	Incidence of acute care complications in vertebral column fracture patients with and without spinal cord injury	D. J. Fletcher et al.	Spine (Phila Pa 1976)	1995 May

in clinical practice to aid medical decision making, we want to further simplify the work and expand the scenarios in which it can be used. Therefore, in the future we hope to package predictive tools into applications for mobile devices, wearables, or personal computers.

CONCLUSION

Our study found that age, OASIS score, SAPS II score, RR, PTT, cardiac arrhythmias, and fluid-electrolyte disorders are predictors of mortality during the ICU stay of thoracic fracture patients without neurological compromise. A multiple logistic regression model and a nomogram were developed and validated. During clinical practice, this nomogram could help physicians screen high-risk patients, make optimal use of resources, and decrease the occurrence of death in this patient population.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HW collected the data, analyzed the data, and drafted the manuscript. TF, YO, MK, JZ, and RD conceived of the study, participated in its design and coordination, and helped to draft the manuscript. YQ was responsible for the whole project, reviewed the manuscript, designed the study, and supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Norman Bethune Program of Jilin University (JJKH20201101KJ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.818439/full#supplementary-material

REFERENCES

1. Niemi-Nikkola V, Saijets N, Ylipoussu H, Kinnunen P, Pesälä J, Mäkelä P, Alen M, Kallinen M, Vainionpää A. Traumatic spinal injuries in northern Finland. *Spine.* (2018) 43:e45–e51. doi: 10.1097/BRS.0000000000002214
2. Pirouzmand F. Epidemiological trends of spine and spinal cord injuries in the largest Canadian adult trauma center from 1986 to 2006. *J Neurosurg Spine.* (2010) 12:131–40. doi: 10.3171/2009.9.SPINE0943
3. Liu P, Yao Y, Liu MY, Fan WL, Chao R, Wang ZG, Liu YC, Zhou JH, Zhao JH. Spinal trauma in mainland China from 2001 to 2007: an epidemiological study based on a nationwide database. *Spine.* (2012) 37:1310–5. doi: 10.1097/BRS.0b013e3182474d8b
4. Doud AN, Weaver AA, Talton JW, Barnard RT, Meredith JW, Stitzel JD, et al. Has the incidence of thoracolumbar spine injuries increased in the United States from 1998 to 2011? *Clin Orthop Relat Res.* (2015) 473:297–304. doi: 10.1007/s11999-014-3870-9
5. Post RB, van der Sluis CK, Leferink VJ, Dijkstra PU, ten Duis HJ. Nonoperatively treated type A spinal fractures: mid-term versus long-term functional outcome. *Int Orthop.* (2009) 33:1055–60. doi: 10.1007/s00264-008-0593-0
6. Defino HL, Canto FR. Low thoracic and lumbar burst fractures: radiographic and functional outcomes. *Eur Spine J.* (2007) 16:1934–43. doi: 10.1007/s00586-007-0406-y
7. Davidson GH, Hamlat CA, Rivara FP, Koepsell TD, Jurkovich GJ, Arbabi S. Long-term survival of adult trauma patients. *Jama.* (2011) 305:1001–7. doi: 10.1001/jama.2011.259
8. Bliuc D, Nguyen ND, Milch VE, Nguyen TV, Eisman JA, Center JR. Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. *Jama.* (2009) 301:513–21. doi: 10.1001/jama.2009.50
9. Center JR, Nguyen TV, Schneider D, Sambrook PN, Eisman JA. Mortality after all major types of osteoporotic fracture in men and women: an observational study. *Lancet.* (1999) 353:878–82. doi: 10.1016/S0140-6736(98)09075-8
10. Niemi-Nikkola V, Saijets N, Ylipoussu H, Kinnunen P, Pesala J, Makela P, et al. Long-term posttraumatic survival of spinal fracture patients in northern Finland. *Spine.* (2018) 43:1657–63. doi: 10.1097/BRS.0000000000002687
11. Darbandsar Mazandarani P, Heydari K, Hatamabadi H, Kashani P, Jamali Danesh Y. Acute physiology and chronic health evaluation (APACHE) III score compared to trauma-injury severity score (TRISS) in predicting mortality of trauma patients. *Emerg.* (2016) 4:88–91.
12. van Breugel JMM, Niemeyer MJS, Houwert RM, Groenwold RHH, Leenen LPH, van Wessem KJP. Global changes in mortality rates in polytrauma patients admitted to the ICU-a systematic review. *World J Emerg Surg.* (2020) 15:55. doi: 10.1186/s13017-020-00330-3
13. Lalwani S, Punia P, Mathur P, Trikha V, Satyarthee G, Misra MC. Hospital acquired infections: preventable cause of mortality in spinal cord injury patients. *J Lab Physicians.* (2014) 6:36–9. doi: 10.4103/0974-2727.129089
14. Krassioukov AV, Furlan JC, Fehlings MG. Medical co-morbidities, secondary complications, and mortality in elderly with acute spinal cord injury. *J Neurotrauma.* (2003) 20:391–9. doi: 10.1089/089771503765172345
15. Ju J, Zhang P, Wang Y, Kou Y, Fu Z, Jiang B, et al. A clinical nomogram predicting unplanned intensive care unit admission after hip fracture surgery. *Surgery.* (2021) 170:291–7. doi: 10.1016/j.surg.2021.01.009
16. O'Connell KM, Quistberg DA, Tessler R, Robinson BRH, Cuschieri J, Maier RV, et al. Decreased risk of delirium with use of regional analgesia in geriatric trauma patients with multiple rib fractures. *Ann Surg.* (2018) 268:534–40. doi: 10.1097/SLA.0000000000002929
17. Tignanelli CJ, Rix A, Napolitano LM, Hemmila MR, Ma S, Kummerfeld E. Association between adherence to evidence-based practices for treatment of patients with traumatic rib fractures and mortality rates among US trauma centers. *JAMA Netw Open.* (2020) 3:e201316. doi: 10.1001/jamanetworkopen.2020.1316
18. Fan T, Wang H, Wang J, Wang W, Guan H, Zhang C. Nomogram to predict the risk of acute kidney injury in patients with diabetic ketoacidosis: an analysis of the MIMIC-III database. *BMC Endocr Disord.* (2021) 21:37. doi: 10.1186/s12902-021-00696-8
19. Tang Z, Chen Y, Ye S, Hu R, Wang H, He J, et al. Fully memristive spiking-neuron learning framework and its applications on pattern recognition and edge detection. *Neurocomputing.* (2020) 403:80–7. doi: 10.1016/j.neucom.2020.04.012
20. Tang Z, Zhu R, Hu R, Chen Y, Wu EQ, Wang H, et al. A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors. *IEEE Trans Cognit Dev Syst.* (2020). 13:645–56. doi: 10.1109/TCDS.2020.3003377
21. Wu EQ, Hu D, Deng P-Y, Tang Z, Cao Y, Zhang W-M, et al. Nonparametric bayesian prior inducing deep network for automatic detection of cognitive status. *IEEE Trans Cybernetics.* (2020) 51:5483–96. doi: 10.1109/TCYB.2020.2977267
22. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
23. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj.* (2015) 350:g7594. doi: 10.1136/bmj.g7594
24. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. *BMC Med Inform Decis Mak.* (2016) 16(Suppl 3):74. doi: 10.1186/s12911-016-0318-z
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33:1–22. doi: 10.18637/jss.v033.i01
26. Hu R, Zhou S, Liu Y, Tang Z. Margin-based Pareto ensemble pruning: An ensemble pruning algorithm that learns to search optimized ensembles. *Comput Intel Neurosci.* (2019) 2019:7560872. doi: 10.1155/2019/7560872
27. Tang Z, Chen Y, Wang Z, Hu R, Wu EQ. Non-spike timing-dependent plasticity learning mechanism for memristive neural networks. *Appl Intelligence.* (2021) 51:3684–95. doi: 10.1007/s10489-020-01985-w
28. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* (1982) 247:2543–6. doi: 10.1001/jama.247.18.2543
29. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* (2008) 8:53. doi: 10.1186/1472-6947-8-53
30. Court-Brown CM, McQueen MM. Global Forum: fractures in the elderly. *J Bone Joint Surg Am.* (2016) 98:e36. doi: 10.2106/JBJS.15.00793
31. Heyde CE, Ertel W, Kayser R. Management of spine injuries in polytraumatized patients. *Orthopade.* (2005) 34:889–905. doi: 10.1007/s00132-005-0847-0
32. Woltmann A, Bühren V. Shock trauma room management of spinal injuries in the framework of multiple trauma. A systematic review of the literature. *Unfallchirurg.* (2004) 107:911–8. doi: 10.1007/s00113-004-0829-4
33. Johansson C, Mellström D, Rosengren K, Rundgren A. Prevalence of vertebral fractures in 85-year-olds. *Radiographic examination of 462 subjects Acta Orthop Scand.* (1993) 64:25–7. doi: 10.3109/17453679308994521
34. Ulvik A, Wentzel-Larsen T, Flaatten H. Trauma patients in the intensive care unit: short- and long-term survival and predictors of 30-day mortality. *Acta Anaesthesiol Scand.* (2007) 51:171–7. doi: 10.1111/j.1399-6576.2006.01207.x
35. Joaquim AF, Ghizoni E, Tedeschi H, Batista UC, Patel AA. Clinical results of patients with thoracolumbar spine trauma treated according to the Thoracolumbar Injury Classification and Severity Score. *J Neurosurg Spine.* (2014) 20:562–7. doi: 10.3171/2014.2.SPINE121114
36. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* (1998) 86:2278–324. doi: 10.1109/5.726791
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2016.* (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
38. Mayer C, Windhager S, Schaefer K, Mitteroecker P. BMI and WHR are reflected in female facial shape and texture: a geometric morphometric image analysis. *PLoS ONE.* (2017) 12:e0169336. doi: 10.1371/journal.pone.0169336
39. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

40. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. Electrocardiogram-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation.* (2021). doi: 10.1161/CIRCULATIONAHA.121.057480
41. Kim H, Lampertico P, Nam J, Lee H, Kim S, Sinn D, et al. An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B. *J Hepatol.* (2021). doi: 10.1016/j.jhep.2021.09.025
42. Moore EE, Moore HB, Kornblith LZ, Neal MD, Hoffman M, Mutch NJ, et al. Trauma-induced coagulopathy. *Nat Rev Dis Primers.* (2021) 7:30. doi: 10.1038/s41572-021-00264-3
43. Fox EE, Holcomb JB, Wade CE, Bulger EM, Tilley BC. Earlier endpoints are required for hemorrhagic shock trials among severely injured patients. *Shock.* (2017) 47:567–73. doi: 10.1097/SHK.0000000000000788
44. Kalkwarf KJ, Drake SA, Yang Y, Thetford C, Myers L, Brock M, et al. Bleeding to death in a big city: an analysis of all trauma deaths from hemorrhage in a metropolitan area during 1 year. *J Trauma Acute Care Surg.* (2020) 89:716–22. doi: 10.1097/TA.0000000000002833
45. Eastridge BJ, Mabry RL, Seguin P, Cantrell J, Tops T, Uribe P, et al. Death on the battlefield (2001-2011): implications for the future of combat casualty care. *J Trauma Acute Care Surg.* (2012) 73(6 Suppl 5):S431–7. doi: 10.1097/TA.0b013e3182755dcc
46. Teixeira PG, Inaba K, Hadjizacharia P, Brown C, Salim A, Rhee P, et al. Preventable or potentially preventable mortality at a mature trauma center. *J Trauma.* (2007) 63:1338–46; discussion 1346–37. doi: 10.1097/TA.0b013e31815078ae
47. Niles SE, McLaughlin DF, Perkins JG, Wade CE, Li Y, Spinella PC, et al. Increased mortality associated with the early coagulopathy of trauma in combat casualties. *J Trauma.* (2008) 64:1459–63; discussion 1463–75. doi: 10.1097/TA.0b013e318174e8bc
48. Johansson PI, Stensballe J, Rasmussen LS, Ostrowski SR. A high admission syndecan-1 level, a marker of endothelial glycocalyx degradation, is associated with inflammation, protein C depletion, fibrinolysis, and increased mortality in trauma patients. *Ann Surg.* (2011) 254:194–200. doi: 10.1097/SLA.0b013e318226113d
49. Cohen MJ, Christie SA. Coagulopathy of trauma. *Crit Care Clin.* (2017) 33:101–18. doi: 10.1016/j.ccc.2016.08.003
50. Davenport RA, Brohi K. Coagulopathy in trauma patients: importance of thrombocyte function? *Curr Opin Anaesthesiol.* (2009) 22:261–6. doi: 10.1097/ACO.0b013e328325a6d9
51. Stansbury LG, Hess AS, Thompson K, Kramer B, Scalea TM, Hess JR. The clinical significance of platelet counts in the first 24 hours after severe injury. *Transfusion.* (2013) 53:783–9. doi: 10.1111/j.1537-2995.2012.03828.x
52. Brown LM, Call MS, Margaret Knudson M, Cohen MJ, Holcomb JB, Wade CE, et al. A normal platelet count may not be enough: the impact of admission platelet count on mortality and transfusion in severely injured trauma patients. *J Trauma.* (2011) 71(2 Suppl 3):S337–42. doi: 10.1097/TA.0b013e318227f67c
53. Kutcher ME, Redick BJ, McCreery RC, Crane IM, Greenberg MD, Cachola LM, et al. Characterization of platelet dysfunction after trauma. *J Trauma Acute Care Surg.* (2012) 73:13–9. doi: 10.1097/TA.0b013e318256deab
54. Goodman S, Weiss Y, Weissman C. Update on cardiac arrhythmias in the ICU. *Curr Opin Crit Care.* (2008) 14:549–54. doi: 10.1097/MCC.0b013e32830a4c5d
55. Agnihotri K, Pothineni NV, Charilaou P, Vaidya VR, Thakkar B, Goyal V, et al. Impact of atrial fibrillation on outcomes with motor vehicle accidents. *Int J Cardiol.* (2018) 250:128–32. doi: 10.1016/j.ijcard.2017.10.002
56. Moss TJ, Calland JF, Enfield KB, Gomez-Manjarres DC, Ruminski C, DiMarco JP, et al. New-onset atrial fibrillation in the critically ill. *Crit Care Med.* (2017) 45:790–7. doi: 10.1097/CCM.0000000000002325
57. Murthy K, Ondrey GJ, Malkani N, Raman G, Hodge MB, Marcantonio AJ, et al. The effects of hyponatremia on bone density and fractures: a systematic review and meta-analysis. *Endocr Pract.* (2019) 25:366–78. doi: 10.4158/EP-2018-0499
58. Schiara LAM, Moirano G, Grosso E, Richiardi L, Tibaldi M, Spertino E, et al. Hyponatremia, hypokalemia, and fragility fractures in old patients: more than an association? *Calcif Tissue Int.* (2020) 106:599–607. doi: 10.1007/s00223-020-00675-6
59. Bhandari SK, Adams AL Li BH, Rhee CM, Sundar S, Krasa H, Danforth KN, et al. *et al*: Sub-acute hyponatraemia more than chronic hyponatraemia is associated with serious falls and hip fractures. *Intern Med J.* (2020) 50:1100–8. doi: 10.1111/imj.14684
60. Sedlacek M, Schoolwerth AC, Remillard BD. Electrolyte disturbances in the intensive care unit. *Semin Dial.* (2006) 19:496–501. doi: 10.1111/j.1525-139X.2006.00212.x
61. Lentsck MH, Oliveira RR, Corona LP, Mathias TAF. Risk factors for death of trauma patients admitted to an Intensive Care Unit. *Rev Lat Am Enfermagem.* (2020) 28:e3236. doi: 10.1590/1518-8345.3482.3236
62. Bokshan SL, Han AL, DePasse JM, Eltorai AE, Marcaccio SE, Palumbo MA, et al. Effect of sarcopenia on postoperative morbidity and mortality after thoracolumbar spine surgery. *Orthopedics.* (2016) 39:e1159–64. doi: 10.3928/01477447-20160811-02
63. Puvanesarajah V, Jain A, Kebaish K, Shaffrey CI, Sciubba DM, De la Garza-Ramos R, et al. Poor nutrition status and lumbar spine fusion surgery in the elderly: readmissions, complications, and mortality. *Spine.* (2017) 42:979–83. doi: 10.1097/BRS.0000000000001969
64. Bruijns SR, Guly HR, Bouamra O, Lecky F, Lee WA. The value of traditional vital signs, shock index, and age-based markers in predicting trauma mortality. *J Trauma Acute Care Surg.* (2013) 74:1432–7. doi: 10.1097/TA.0b013e31829246c7
65. Hardin KS, Leasia KN, Haenel J, Moore EE, Burlew CC, Pieracci FM. The sequential clinical assessment of respiratory function (SCARF) score: a dynamic pulmonary physiologic score that predicts adverse outcomes in critically ill rib fracture patients. *J Trauma Acute Care Surg.* (2019) 87:1260–8. doi: 10.1097/TA.0000000000002480
66. El-Manzalawy Y, Abbas M, Hoaglund I, Cerna AU, Morland TB, Haggerty CM, et al. OASIS +: leveraging machine learning to improve the prognostic accuracy of OASIS severity score for predicting in-hospital mortality. *BMC Med Inform Decis Mak.* (2021) 21:156. doi: 10.1186/s12911-021-01517-7
67. Fueglistaler P, Amsler F, Schuepp M, Fueglistaler-Montali I, Attenberger C, Pargger H, et al. Prognostic value of Sequential Organ Failure Assessment and Simplified Acute Physiology II Score compared with trauma scores in the outcome of multiple-trauma patients. *Am J Surg.* (2010) 200:204–14. doi: 10.1016/j.amjsurg.2009.08.035

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Ou, Fan, Zhao, Kang, Dong and Qu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fpubh.2021.821410

Hybrid Deep Neural Network for Handling Data Imbalance in Precursor MicroRNA

Elakkiya R.[1], Deepak Kumar Jain[2], Ketan Kotecha[3], Sharnil Pandya[4], Sai Siddhartha Reddy[1], Rajalakshmi E.[1], Vijayakumar Varadarajan[5], Aniket Mahanti[6] and Subramaniyaswamy V[1]*

[1] School of Computing, SASTRA Deemed University, Thanjavur, India, [2] College of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China, [3] Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India, [4] Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India, [5] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia, [6] The University of Auckland, Auckland, New Zealand

Over the last decade, the field of bioinformatics has been increasing rapidly. Robust bioinformatics tools are going to play a vital role in future progress. Scientists working in the field of bioinformatics conduct a large number of researches to extract knowledge from the biological data available. Several bioinformatics issues have evolved as a result of the creation of massive amounts of unbalanced data. The classification of precursor microRNA (pre miRNA) from the imbalanced RNA genome data is one such problem. The examinations proved that pre miRNAs (precursor microRNAs) could serve as oncogene or tumor suppressors in various cancer types. This paper introduces a Hybrid Deep Neural Network framework (H-DNN) for the classification of pre miRNA in imbalanced data. The proposed H-DNN framework is an integration of Deep Artificial Neural Networks (Deep ANN) and Deep Decision Tree Classifiers. The Deep ANN in the proposed H-DNN helps to extract the meaningful features and the Deep Decision Tree Classifier helps to classify the pre miRNA accurately. Experimentation of H-DNN was done with genomes of animals, plants, humans, and Arabidopsis with an imbalance ratio up to 1:5000 and virus with a ratio of 1:400. Experimental results showed an accuracy of more than 99% in all the cases and the time complexity of the proposed H-DNN is also very less when compared with the other existing approaches.

Keywords: bioinformatics, precursor microRNA, hybrid deep neural network, deep artificial neural network, deep decision tree classifier

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Prabaharan Sengodan, Malla Reddy Institute of Engineering & Technology, India
Devi Mani, King Khalid University, Saudi Arabia

***Correspondence:**
Ketan Kotecha
head@scaai.siu.edu.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *24 November 2021*
Accepted: *03 December 2021*
Published: *23 December 2021*

Citation:
R. E, Jain DK, Kotecha K, Pandya S, Reddy SS, E. R, Varadarajan V, Mahanti A and V S (2021) Hybrid Deep Neural Network for Handling Data Imbalance in Precursor MicroRNA. Front. Public Health 9:821410. doi: 10.3389/fpubh.2021.821410

INTRODUCTION

The problem of imbalanced data has often been recognized as a significant challenge in Machine and Deep Learning. The ML classifiers tend to get influenced by the majority classes when it comes to imbalanced data. This results in a poor classification of the minority classes. Whenever the class sizes are significantly different, the classification methods perform much better in the larger class, achieving accurate results, whereas the minority class seems to have a very poor recall. The classification of one class being considerably neglected in comparison to another is still one of the most difficult problems to solve in the creation of new classification models today. Bioinformatics is one of the major streams that provide a large number of highly imbalanced data. These enormous

datasets require a great solution that can handle such problems and are appropriate for such imbalance ratios. One of the major data imbalance problems in bioinformatics is the identification of the precursor micro RNAs (pre-miRNA) from the genome. The miRNA is a special type of non-coding RNA. They have a length of 21 nucleotides and they play a key role in the posttranscriptional regulation of gene expression. The miRNA is a minute, endogenous RNA having a hairpin-like structure. Numerous key biological processes, such as disease genesis and progression, are regulated by miRNAs. They can act as an oncogene or tumor suppressor in a variety of cancers assisting in prognosis prediction, therapeutic assessment, and disease diagnosis (1). A promising therapy for breast cancer is delivered by the latest miRNA-based drugs (2).

The genome data consists of hundreds of thousands of sequences out of which there will be only a few miRNAs. Moreover, there are several other similar RNAs which makes it even more difficult to differentiate them. Further, the minority class usually contains a small number of high-variability instances, making it more challenging for a classifier the generalization of the new dataset. So, to find them in the highly imbalanced genome data and classify them is a great challenge and hence there is a need to find an optimal solution for identifying them with a simple procedure. Although the computational methods have been necessary tools in miRNA gene finding and functional studies (3). The Machine Learning (ML) techniques have made it possible to classify them to a good extent (4). Machine Learning and Deep Learning (DL) have taken control over every field and are providing adequate solutions for every complex problem. There are few algorithms proposed for which the True Positive Rate (TPR) is very less and complexity is more. For the genome datasets, there are many negative samples and very few positive samples of miRNA sequences. Generally, this type of problem is related to Imbalanced data. To handle such Imbalanced data there are a few solutions out of which Sampling is the most famous and efficient way. There are two variants of sampling: Oversampling and Undersampling. Oversampling is the technique in which the fewer samples in the dataset will be upscale (in other words upscaling the minorities) to make the data to be balanced. In this technique, many dummies for fewer samples are generated. When the dataset is with a smaller number of samples (both positive and negative) this technique will help to build a model with an adequate number of samples. Under Sampling is the technique in which the greater number of samples in the dataset will be downscaled (in other words downscaling the majorities) to make the data to be balanced. In this technique, the Selection method or Random sample method is chosen. In general, when the dataset is very large, the model will take a huge time to execute, this technique will help us to overcome this issue. In one of the proposed models (5), they chose an Imbalance Ratio (IR) up to 1:2000 (meaning one positive sample and 2,000 negative samples), and different ratios like 1:1, 1:100, 1:500, 1:1000, 1:1500, 1:2000. To improve the model and follow the consistency, different IRs like 1:2500, 1:3000, 1:3500, 1:4000, 1:4500, 1:5000 have been added. A model was proposed where, in addition to the Animal and Plant dataset, the Human and Arabidopsis dataset with IR up to 1:5000 and virus dataset with IR up to 1:400 was considered to enhance the performance (6).

The existing systems like Multi-layer perceptron (MLP), Support Vector Machine (SVM), Deep Self-Organizing Map (DeepSOM), Deep Elastic SOM(DeSOM), Deep Elastic Ensemble SOM (DeeSOM), Deep Belief Neural Network (DeepBN) can solve this problem to some extent only. The Deep Architectures which work without SMOTE (Synthetic Minority Oversampling Technique) (7) take more time and accuracy is also poor, the other systems stated above also produce low accuracy if the class imbalance ratio is gradually increased. Most of the existing works so far have not considered multiclass pre-miRNA classification having an imbalanced dataset, even if they have discussed their performance is not up to the mark and hence end up having low accuracy. The accuracy of the existing work also goes down as the imbalance ratio is increased as it is given in the real-time scenarios. So, to address the problems like high run time and low accuracy faced in these existing systems, we propose a Hybrid Deep Neural Network (H-DNN) architecture, which is implemented by integrating the Deep Artificial Neural Networks (ANN) with a Deep Decision Tree classifier (Deep DT). Deep ANN is capable of optimizing a wide range of coefficients. As a result, Deep ANN can tolerate a much larger variation than traditional models. With enough training, Deep ANN will be prepared for any circumstance. It is a type of machine learning technology with a large memory capacity since it is capable of remembering every call. On imbalanced data, DT usually performs effectively. They operate by understanding a hierarchy of if/else scenarios, which forces them to address both majorities as well as minority classes. The main contributions of the proposed work can be listed as follows:

- Preprocessing of multiclass genome dataset and selection of features using Select-K-Best and Recursive Feature Elimination.
- Building proposed H-DNN framework by integrating Deep Artificial Neural Network and Deep Decision Tree for classification of miRNA from a multiclass genome dataset.
- Apply the H-DNN architecture with different imbalance ratios of multiclass genome dataset consisting of Animal, Plant, Arabidopsis, Human, and Virus genomes.

The rest of the paper is organized as follows: Section Related Work discusses the related work. Section Proposed Work discusses the proposed H-DNN architecture and its components. Section Experimental Work details the experiments and presents the results of H-DNN and section Results and Discussions summarizes the proposed approach and its future enhancements.

RELATED WORK

Due to the imbalanced class distributions in the data, handling such class imbalance problems and drawing a large amount of interest toward it, has become the major issue. To deal with these class imbalance problems many ensemble methods are proposed but most of them focus on the two-class imbalance problems, leaving many issues unsolved in the multiclass imbalance

problems. To solve those multi-class problems efficiently an algorithm called AdaBoost.NC was proposed. It produced better results when compared with other popular ensemble Techniques (8). As the availability is expanding in all kinds of fields it becomes difficult to perform analysis and progress from raw data to decision-making processes. Even though new data engineering techniques are showing great success in solving real-world problems, learning from imbalanced data has emerged as a new challenge. To enable efficient learning from imbalanced data it is required to have new approaches, understandings, algorithms, principles, and tools, because of its complex characteristics (9, 10). Class imbalances make supervised learning problems difficult because some classes have more examples than others. The existing methods have completely focused on binary-class cases. A dynamic sampling method for Multilayer Perceptron (MLP) was proposed to solve the multi-class imbalance problems. To train the MLP, DyS dynamically selects informative data. This Dynamic sampling technique has performed well when compared with other techniques (11). There are several other preprocessing techniques for the data, classification algorithms, and evaluating the model involved in modeling methods (12). The important regulators of gene expression are MicroRNAs (miRNAs). The genetic loss of the tumor suppressor miRNAs or Amplification and the overexpression of one's 'oncomiRs' is related to human cancer and is enough to cause tumorigenesis in mouse models. Besides, the depletion of global miRNA by genetics and in the components of miRNA biogenesis machinery epigenetic alterations is oncogenic. All these together with novel miRNA regulatory pathways and factors show how important miRNA dysregulation is in cancer (13). Even after identifying hundreds of miRNAs in different species, a lot of others remain unknown. So, in understanding miRNA-mediated posttranscriptional regulation mechanisms the discovery of new miRNA genes is an important step. The biological approaches in identifying miRNA genes might be limited. So, to overcome those limitations in the biological approaches sophisticated computational approaches are followed to identify the possible miRNA (14). Some of the work deals with the miRNAs in terms of diagnosing dreadful diseases in the medical field. Several web-based bioinformatics tools were proposed for the analysis of miRNA such as expression, detection, etc. with an easy-to-use model that doesn't require any previous knowledge of utilization (15). A Support Vector Machine (SVM) based approach was proposed to classify the real and pseudo miRNA. It achieved an accuracy of up to 90% (16). The model was exclusively for the human dataset but the remaining datasets could also be used. In this work, then genomic information was not required. Later, software that classified the RNA of mammals, nematodes, and also urochordates was proposed. A new methodology was proposed based on Support Vector Machine (SVM) and named

TABLE 1 | Comparison of various state-of-the-art methodologies.

Author	Methodology used	Metrics	Dataset
Park (25)	Recurrent Neural Network (RNN).	Various Metrics were used. Experimental results yielded an F1-score of 0.93 for the Human dataset, 0.94 for the Cross-Species dataset, and 0.93 for the New pre-miRNA dataset.	Human, Cross-Species and New miRNA
Thomas et al. (26)	Deep Learning with a restricted Boltzmann machine (RBM) for classification and Modified sampling technique is applied to address the class imbalance problem.	Various Metrics were used. The model gave an accuracy of 0.968	Human Dataset
Stegmayer et al. (27)	Deep Self-organizing maps (DeepSOM) with Clustering	Various evaluation metrics were used. The model provided almost 95% of Gm and 97% of accuracy for the most imbalanced data.	H. sapien and A. thaliana, Animal, and Plant
Bugnon et al. (5)	Hybrid deep learning method with CNN and LSTM.	Various Metrics were used for different imbalance ratios. Experimental results yielded an F1-score of more than 40% for Animal and more than 33% for the Plant dataset.	Animal and Plant
Tang et al. (28)	CNN with different types of feature learning and encoding methods.	The results gave an accuracy of 99.25% and an F1-score of 99.25% for Rfam-300	Rfam-300, Rfam-120, Rfam-60, Rfam-30
Shi et al. (29)	Localized multiple kernel learning model with a nonlinear synthetic kernel (LMKL-D)	The model gave an accuracy of 98.3%, sensitivity of 93.06 Specificity of 99.27 and mean of 96.11.	pre-microRNAsand pseudo pre-microRNA
Yones et al. (30)	Convolutional deep residual neural network	The model provided a precision of 0.93 for A. thaliana full genome, 0.67 for A. gambiae, and 0.50 for H. Sapiens	A. thaliana, C. elegans, A. gambiae, and H. Sapiens
Proposed HDNN Methodology	Artificial Neural Network embedded with Decision Tree	Various metrics were used for evaluation for different imbalance ratios. The results yielded F1 Score of more than 0.50 (0.50–0.93) for Animal, more than 0.58 (0.58–0.95) for Plant, more than 0.60 (0.60–0.94) for Human, more than 0.50 (0.50–0.95) for Arabidopsis and more than 0.50 (0.50–0.94) for Virus.	Animal, Plant, Human, Arabidopsis, and Virus.

as Microprocessor SVM, which predicted with 50% accuracy (17). To improve the accuracy of this model, the model was trained with another classifier and ensemble both to get better results than acquire 90% in total. Another work dealt with the miRNA sequences of the mouse and human using the SVM approach (18). The model worked with around 377 mouse data and 475 human data and with the candidates, the dataset had 3,441 humans and 3,476 mouse samples in total. The classification was done according to the structure of the miRNAs, the model names, and the mirCoS. One of the works also dealt with the use of ab initio Prediction model clustered microRNA identification (19).

An application for the classification of pre-miRNAs was proposed which dealt with the classification of the mammalian miRNAs such as Human, Rat, and Mouse (20). This approach analyzed the genome properly using the model named ab initio which gives an accuracy of up to 60% when they undergo cross-validation. To analyze the structure of miRNA, a tool named MiRFinder was proposed which was built with the SVM model (21). The structure of the miRNA is a hairpin-like structure, the prediction mainly focused on the structure and later the features were used to predict the miRNA. This tool was mainly used for the genome-wide search of the miRNAs. The dataset which has 697 real miRNA and 5,428 pseudos from different sources was collected and experimented with. The tool gained sensitivity up to 93%. Another model named miRenSVM was proposed that gave good results for the data in miRBase 13.0 animal and other species (22). For animal data in miRBase 13.0, it gave 100% sensitivity. A novel RNA effective classification methodology namely microPred was proposed which classified the types of miRNA using the Machine Learning approach (23). It used a classifier namely microPred, which also could handle the data imbalance issue as well as new feature extraction relevant biologically. It achieved a specificity of 97.28% and a sensitivity of 90.02%. The model gave the prediction rate of 92.71% for humans and 94.24% for the virus. Another new model was developed that used a method named YamiPred, which is an ensemble of Genetic Algorithms (GA) and Support Vector Machine (SVM) for the prediction of miRNAs (24). The model mainly used a human dataset to predict the human miRNAs. The model appeared to be robust and efficient with different performance characteristics. **Table 1** illustrates the comparison of various state-of-the-art methodologies used for the classification of miRNA.

PROPOSED WORK

In this section, we discuss the proposed Hybrid Deep Neural Network approach for the classification of miRNA from genome data. Different genome datasets have been considered in our proposed work such as Animal, Plant, Human, Arabidopsis, and Virus datasets. First, the dataset is preprocessed and all

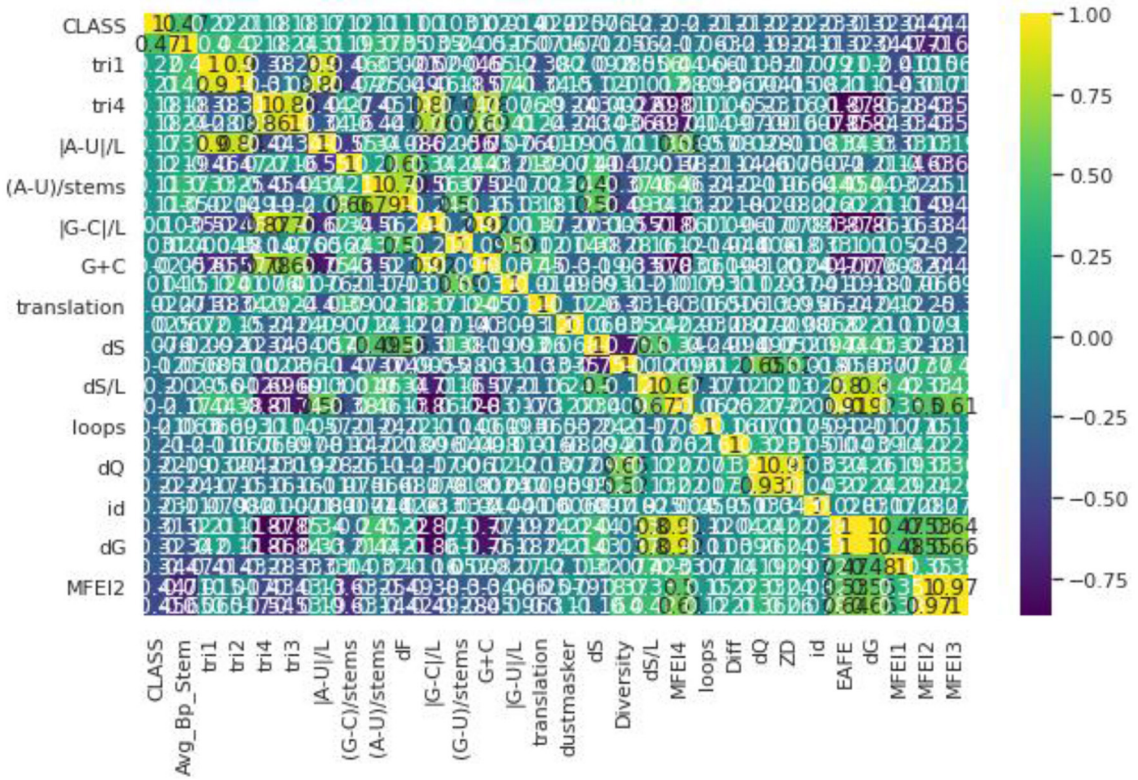


FIGURE 1 | Correlation matrix.

the important features are selected and extracted and the non-useful features are omitted. For feature selection and extraction feature selection Algorithms like Select-K-Best and Recursive Feature Elimination are used. After preprocessing the dataset, the proposed H-DNN algorithm is implemented. The proposed H-DNN framework is the integration of Deep ANN with the Deep DT. Then the implemented model has experimented with the different genome datasets (31).

Preprocessing

Pre-processing refers to Feature Selection or Feature Extraction. Different Feature Selection algorithms have been applied with the dataset (32). To have a clear idea of feature selection, the dataset has to be analyzed first. Different genome datasets have been considered in our proposed work such as Animal, Plant, Human, Arabidopsis, and Virus datasets. Even though these datasets have a different number of samples in each, they have the same number of features (33). The dataset used here has a total of 30 features in which one is a serial number, hence we omit that attribute. The remaining 29 features are taken into consideration and passed to the Feature Selection Algorithms (34). **Figure 1** illustrates the correlation matrix plotted for all the features and a heat map is produced out of that as shown. The correlation matrix helps us to characterize and sum up the data for our understanding. Correlation, here, helps us to compute the association between two variables by testing the effectiveness of their linear relationship. After estimating the correlation matrix, the Feature selection Algorithms like Select-K-Best and Recursive Feature Elimination are used. In the Select-K-Best algorithm, the features are assigned with particular scores. The Selection is generally made based on the scores which are greater than the mean of all the scores in this Feature Selection (35). In the Recursive Feature Elimination, the rank is assigned to each feature and the ranks may be the same for more than one feature. The feature selection is made based on the ranks that are assigned. Generally, the selection is done based on the first few ranks like 10, 12, etc., based on several features (36). The Recursive Feature Elimination with Cross-Validation is used. This gives scores between 0 and 1. The 0, being the least important and one is the most important feature to build any model. This algorithm is applied to the dataset and the results state that all the features are nearer to the one (37). Hence, all the features are important. This proves that there is no need for removal or elimination of any feature from the data so all 29 features are used to build the proposed model (38).

Proposed H-DNN Architecture

In this section, we discuss our proposed work. We have developed a Hybrid Deep Neural Network (H-DNN) wherein we have integrated Deep Artificial Neural Network and Deep Decision Tree (DT) classification technique for classifying the miRNAs in the genome data. The integration of feature extraction using machine learning techniques and feature analysis and classification using Deep Neural Network has provided better results. The experimental results have been carried out using Animal, Plant, Human, Arabidopsis and, Virus genome datasets. The results are then analyzed and compared for the model using Deep ANN and Deep DT. Deep ANN here helps in analyzing the features by optimizing a wide range of coefficients. As a result, Deep ANN can tolerate a much larger variation than traditional models. Deep ANN gets well prepared for any circumstance with sufficient training. It is a type of machine learning technology with a large memory capacity since it is capable of remembering every call. DT usually performs effectively in case of imbalanced data by understanding a hierarchy of making decisions for different scenarios, which forces them to address both majorities as well as minority classes.

The structure of ANN consists of a neuron, a propagation function, and a bias. Consider a neuron k having an input $I_k(t)$ received from the previously connected neurons. The neuron will have an activation function $a_k(t)$, an optional threshold θ_k, an activation function f for computing new activation for time $t+1$ as given in Equation 1, and a function f_{out} for computing the activation output which is formulated as given in Equation 2.

$$a_k(t+1) = f(a_k(t), p_k(t), \theta_k) \quad (1)$$

$$O_k = f_{out}(a_k(t)) \quad (2)$$

The input, $I_m(t)$ to the neuron m from the output of k, is computed using the propagation function defined as given by Equation 3. After adding the bias, w_{0m}, we formulate the input as given in Equation 4.

$$I_m(t) = \sum_k O_k(t) w_{km} \quad (3)$$

$$I_m(t) = \sum_k O_k(t) w_{km} + w_{0m} \quad (4)$$

Artificial Neural Networks (ANNs) have several different coefficients, which the ANNs can optimize. So, when compared with the traditional models, ANNs can handle much more

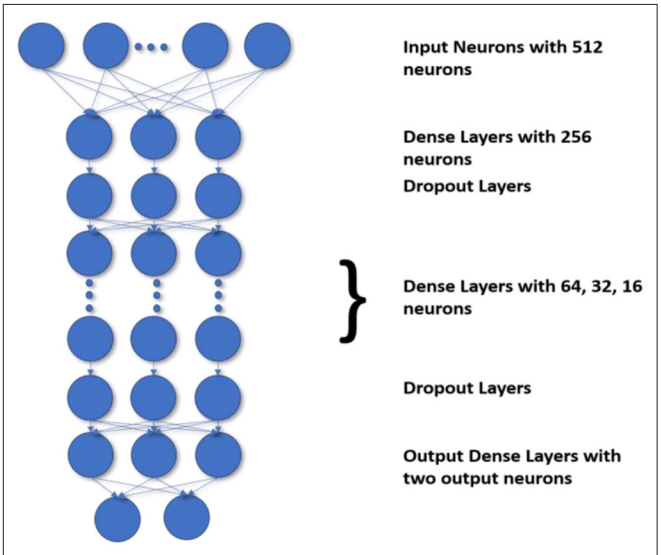


FIGURE 2 | Structure of ANN in each node of Decision Tree in the proposed framework.

variability (39). This makes ANN a stronger model when it comes to memorization (40). A decision tree is a supervised learning Algorithm. Both regression and classification problems can be solved using Decision Trees. In the tree representation, each leaf node represents a class label and the internal node of the tree represents the attributes. Decision trees include three main units that are node, branch, and leaf. The node is the decision. The branch is the potential decision, and the leaf is the potential outcome of each decision. The DT illustrates the long-term consequences of certain decisions. The trade-offs and probabilities can also be expressed using decision trees. In a decision tree, after every level, we have a split taking place which in turn results in the next level. Therefore, the major challenge in Decision trees is to identify the attribute for the root node in each level. This is called attribute selection. There are various popular attribute selection measures in decision trees we have used Gini Index for attribute selection in our framework. Gini index can be formulated as shown in Equation 5.

$$Gini\ Index\ =\ 1 - \sum_{j} p_j^2 \tag{5}$$

here p_j denoted probability for class j. Gini index is used here as a cost function for the split evaluation in feature selection. When compared with other algorithms, the decision tree needs less effort for data preparation during preprocessing. It does not require scaling of data or normalization of the data. In addition to this transparency and ease of use makes the decision tree a better model to handle large amounts of data.

The most challenging task is to handle the imbalanced data. The current state-of-the-art methods do not define the decision process take more time and accuracy is also poor. Generally, the state-of-the-art frameworks produce low accuracy if the class imbalance ratio is gradually increased. Before DNN, decision trees were the standard method used to improve accuracy and performance. Though as compared to DNN the accuracy gained by using a decision tree for classification are less the DT preserves the interpretability. So to preserve the DT's interpretability and match the DNN's performance, we use the H-DNN architecture wherein we integrate the Decision Tree and the ANN. The H-DNN model consists of a DT wherein each node of the DT will have an ANN model. The DT is used to make the classification of miRNA in an H-DNN, maintaining high-level interpretability. Every node in a decision tree, on the other hand, is a neural network that makes low-level decisions. First, a hierarchy for DT is constructed and then the sample is passed through the ANN, which is a fully connected neural network, in each decision node. Then the model computes the analysis by executing the final fully connected layer as a series of integrated decision rules. The final prediction is a culmination of these decisions. The ANN structure proposed in this paper deals with a dense layer as input layer with 512 nodes and ReLu activation function followed by another dense layer with 256 nodes and a dropout layer, then three more dense layers with 64, 32, 16 nodes respectively with ReLu activation function and a dropout layer and finally one dense layer with two nodes as output layer. This ANN structure, as illustrated in **Figure 2**, is embedded into each node of the decision tree for making the decision rules. In the fully connected layer of each node of the decision tree, the inner product of the matrix-vector is calculated and an inner product that constitutes the maximum value is chosen for the class decision rule as shown in Equation 6.

Algorithm 1 Proposed H-DNN.

Input: Genome data, D, with attributes, a
Output: Decision Tree with classes

1. $tree = \{\ \}$
2. $minVal = 0$
3. **for** *each* $a \in D$ **do** :
   //Considering Input Layer of ANN
4. **for** *each node* $n \in a$ *and input* x_n **do** :
5. $out_n = x_n$
   //Considering Hidden Layer of ANN
6. **for** *each hidden node* h **do** :
7. $out_h = \sum_n \varnothing(w_{hn}.out_n)$
8. $g = GiniIndex(out_h,\ d)$
9. **for** *each output node* k **do** :
10. $out_k = \sum_k \varnothing(w_{kh}.out_h)$
11. *Train ANN using Backpropagation*(out_h)
12. **if** $g < minVal$ **then**
13. $minVal = g$
14. $tree' = \{a\}$
15. $partition(tree,\ tree')$
16. ***repeat*** till all the partitions are processed
17. ***return*** tree

$$[w_1\ w_2 \vdots w_n\ ][\vdots x \vdots\ ] = [(x, w_1)\ (x, w_2) \vdots (x, w_n)\ ]$$
$$= [y_1\ y_2 \vdots y_n\ ] \rightarrow argmax(y) \tag{6}$$

Algorithm 1 illustrates the H-DNN algorithm that uses the Deep Decision Tree having Artificial Neural Network embedded in each root node. The ANN framework in the proposed model has a single input layer, a few hidden layers, and a single output layer. In the final layer of the hidden layer, the decision rule for splitting the tree has been incorporated using Gini Index. Line 1 and 2 in **Algorithm 1** initialize the tree and the minimum criteria to be considered for making the decision rule. The Deep ANN is embedded into each node of the Deep DT as illustrated from lines 3 to 10. For a given input node and input, line 4 and 5 illustrates the implementation of the first layer, i.e., the input layer of the Deep ANN. Using the output of the input layer the hidden layer is calculated using lines 6 and 7. In line 8 the Gini Index is calculated to determine the decision rule for splitting the node or tree further. Using the output of the hidden layers the output layer is constructed in lines 9 and 10. In line 11 the Backpropagation algorithm is called to adjust the weight according to the errors. In lines 12 to 15, the decision rule has been implemented. Once the decision rule is applied the nodes are split accordingly as shown in line 16. This process is then implemented for each node until all the nodes are processed and no further nodes can be split.

Algorithm 2 illustrates the implementation of the backpropagation algorithm used for generalizing the Neural Network (NN) by fine-tuning the weights based on the previous iteration's error rate. After applying the forward propagation, the backpropagation algorithm is called to fine-tune the NN. From lines 1 to 6 in **Algorithm 2** all the required parameters are initialized. From lines 7 to 13 the backpropagation process is applied to the neural network.

The proposed architecture can be divided into four modules as illustrated in **Figure 3**. Module one is the Imbalanced genome dataset which has the collection of Plant, Animal, Human, Arabidopsis, and Virus genome data. This data is preprocessed in module two wherein the feature variables are correlated to find the relationship between the features. In this module, the required features are extracted using the Select-K-Best Method and Recursive Feature Elimination Method, and the rest of the unwanted attributes are discarded. After the feature selection and extraction, the preprocessed data is passed to the third module. The third module is the proposed H-DNN architecture. The H-DNN architecture is the integration of the Decision Tree and the ANN framework. The ANN framework used here helps to avoid overfitting and thereby increasing the performance. The Decision tree helps to make a decision and examine the benefits and cons of each final option. Every node in a decision tree, on the other hand, is a neural network that makes low-level decisions. The inputs of each node of the Decision Tree are passed to ANN for making the decision and analyzing the probability values of the class. The final evaluation of the model is done in the fourth Module, the Model Evaluation module, using Accuracy, Specificity, Precision, and F1-score as the evaluation metrics.

Algorithm 2 Modified Backpropagation(a).

1. $d \leftarrow$ *training dataset mxn dimension*
2. $y \leftarrow$ *class labels*
3. $w \leftarrow$ *weights*
4. $l \leftarrow$ *number of layers* $1, \ldots .., z$
5. $error_{ij}^{l} \leftarrow$ *error for each* i, j, l
6. $b_{ij}^{l} \leftarrow 0$ *for each* i, j, l
7. ***for*** $i = 1$ *to* m **do** :
8. $d^{l} \leftarrow a(z) - y(i)$
9. $b_{ij}^{l} \leftarrow b_{ij}^{l} + a_{j}^{l}.b_{i}^{l+1}$
10. ***if*** $j \neq 0$ **then**
11. $error_{ij}^{l} \leftarrow \frac{1}{m} b_{ij}^{l} + \lambda w_{ij}^{l}$
12. ***else***
13. $error_{ij}^{l} \leftarrow \frac{1}{m} b_{ij}^{l}$

EXPERIMENTAL RESULTS

Experimental Work

In this section, the dataset and the experimental results will be explained in detail by each dataset (41). In the Animal dataset, there are 7,053 positive and 218,154 negative samples and in total there are 225,207 animal sequences. One lakh seventeen thousands one hundred and one plant sequences are taken in which 2,154 positive and 114,929 negative samples are present. One thousand four hundred six positives and 81,228 negatives

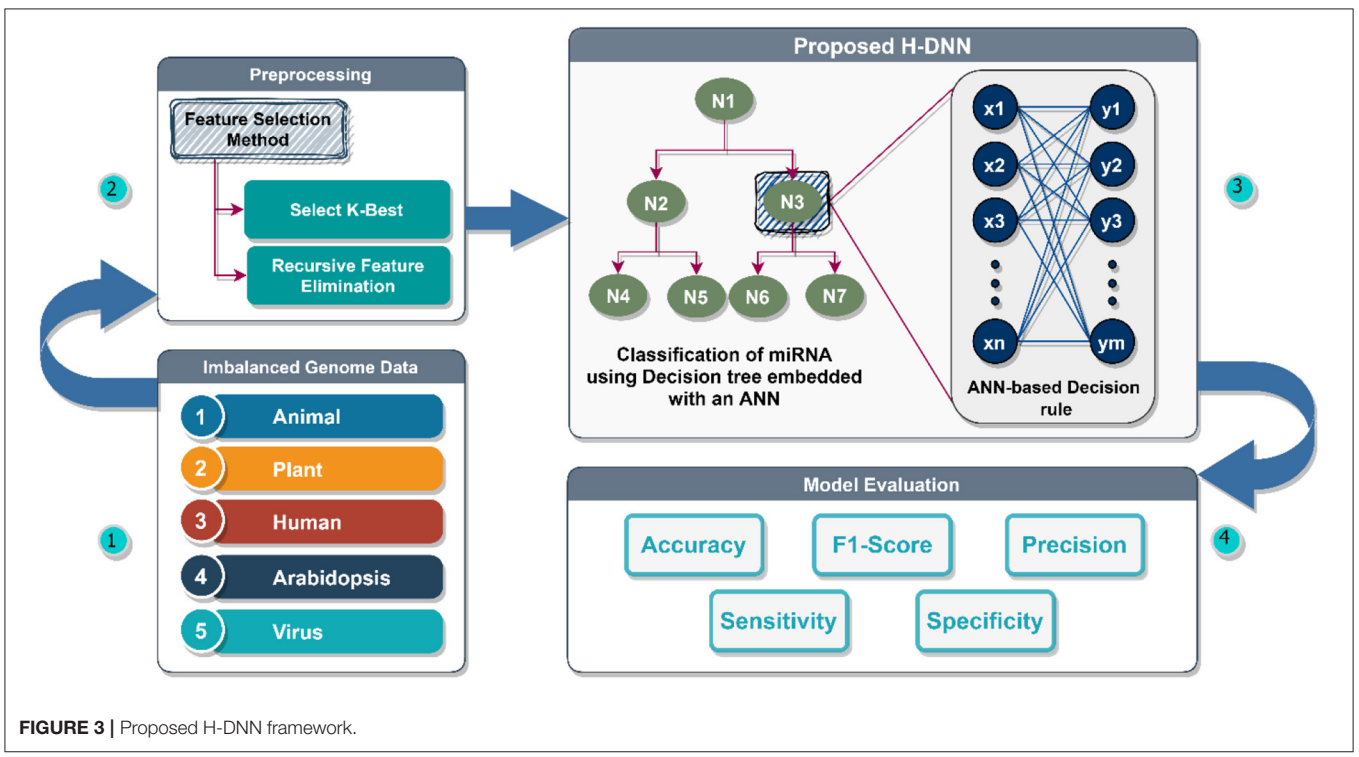


FIGURE 3 | Proposed H-DNN framework.

TABLE 2 | The no. of positive and negative samples for different Imbalance Ratios.

IR	Animals		Plants		Human		Arabidopsis	
	No. of positives	No. of negatives	No. of positives	No. of negatives	No. of positives	No. of negatives	No. of positives	No. of negatives
1:1	7,053	7,053	2,172	2,172	1,406	1,406	231	231
1:100	2,182	218,154	1,149	114,929	812	81,228	231	23,100
1:500	436	218,154	230	114,929	162	81,228	57	28,359
1:1000	218	218,154	115	114,929	81	81,228	28	28,359
1:1500	145	218,154	77	114,929	54	81,228	19	28,359
1:2000	109	218,154	57	114,929	41	81,228	14	28,359
1:2500	87	218,154	46	114,929	32	81,228	11	28,359
1:3000	73	218,154	38	114,929	27	81,228	9	28,359
1:3500	62	218,154	33	114,929	23	81,228	8	28,359
1:4000	55	218,154	29	114,929	20	81,228	7	28,359
1:4500	48	218,154	26	114,929	18	81,228	6	28,359
1:5000	44	218,154	23	114,929	16	81,228	6	28,359

TABLE 3 | The number of positive and negative samples for different IRs for virus.

IR	Virus	
	No. of positives	No. of negatives
1:1	237	237
1:50	17	839
1:100	8	839
1:200	4	839
1:300	3	839
1:400	2	839

and in total 82,634 human samples are given in this dataset. In Arabidopsis 231 positive samples and 28,359 negative samples in total, and there are 28,590 sequences. The virus samples are very few so it has only 1,076 in a total of which 237 are positive and 839 are negative samples (26). This is treated as a separate one and the maximum IR is only 1:400 (27). For the Animal, Plant, Human, and Arabidopsis datasets, the number of positive and negative samples for different IRs are tabulated in **Table 2**. For the virus, the number of positive and negative samples for different IRs is listed in **Table 3**. The initial target is to find the specificity and other metrics till IR up to 1:2000. Later, to improve the model it was extended to 1:5000. For the classification, here two Algorithms are used. They are ANN and Decision Tree Classifiers (10, 25, 28, 39, 40, 42–44).

RESULTS AND DISCUSSIONS

As discussed in section Proposed Work the ANN has been integrated with the DT in our proposed work and the results are represented graphically. Graphs are plotted for clear visualization of the results. As explained in the previous sections, the model is improvised by using different datasets like Animals, Plants, Humans, Arabidopsis, and Viruses genomes. The work is extended in terms of the Imbalance Ratios also it is extended up to 1:5000. The model gives some predictions even for the 1:5000 ratio also, this can be visualized individually. **Figure 4** illustrates the graphical representation of experimental results done using the proposed H-DNN framework on Animal, Plant, Human, Arabidopsis and, Virus genome datasets. The x-axis in the graph indicates the IR and the y-axis in the graph plotted illustrates the metric values. The result analysis of the proposed H-DNN Algorithm was done based on the evaluation metrics: Accuracy (Acc), Specificity (SP), Sensitivity (SE), F1-score (F1), and Precision (Prec).

First, the ANN and Decision Tree Classifier results were analyzed individually. Then their results were compared with the proposed H-DNN framework. The Experimental analysis was done by implementing the state-of-the-art ANN and DT separately and then by implementing the proposed H-DNN approach using the above-said dataset. The dataset was trained batch-wise. **Table 4** illustrates the result comparison of the proposed H-DNN Algorithm with the ANN and the DT algorithms for the Animal dataset. **Table 5** illustrates the result comparison of the proposed H-DNN Algorithm with the ANN and the DT algorithms for the Plant dataset.

Table 6 illustrates the Experimental results for the Human dataset. **Table 7** illustrates the Experimental results for the Arabidopsis datasets. **Table 8** illustrates the Experimental results for the Virus datasets.

By analyzing the results, we observed that the ANN gives better results and gives a good True Positive Ratio (TPR) with 2 epochs and up to IR 1:2000. If the model is trained further for more epochs, then we are getting an over-fitted model. The time required is also in milliseconds which is very optimal whereas the existing system takes hours to train one epoch. Furthermore, when the model needs to be trained with IR up to 1:5000 Decision Tree Classifier gives better results with good TPR. But when we integrate both the DT and

TABLE 4 | Experimental results using animal dataset.

	ANN Algorithm					DT Algorithm					Proposed H-DNN Algorithm				
IR	**Acc**	**SP**	**SE**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**
1	0.51	0.03	1.00	0.37	0.73	0.57	0.49	0.96	0.65	0.83	0.93	0.94	0.93	0.93	0.93
100	0.99	0.99	0.66	0.74	0.69	0.99	0.81	0.83	0.78	0.76	0.99	0.64	1.00	0.82	0.82
500	1.00	1.00	0.53	0.50	0.50	1.00	0.72	0.77	0.60	0.59	1.00	0.44	1.00	0.70	0.68
1,000	1.00	1.00	0.67	0.50	0.50	1.00	0.71	0.83	0.59	0.59	1.00	0.41	1.00	0.69	0.67
1,500	1.00	1.00	0.60	0.50	0.50	1.00	0.71	0.80	0.61	0.61	1.00	0.41	1.00	0.72	0.73
2,000	1.00	1.00	0.44	0.50	0.50	1.00	0.63	0.72	0.57	0.58	1.00	0.26	1.00	0.64	0.65
2,500	1.00	1.00	0.45	0.50	0.50	1.00	0.58	0.73	0.54	0.55	1.00	0.15	1.00	0.58	0.59
3,000	1.00	1.00	0.33	0.50	0.50	1.00	0.63	0.67	0.58	0.60	1.00	0.26	1.00	0.66	0.69
3,500	1.00	1.00	0.13	0.50	0.50	1.00	0.60	0.56	0.55	0.55	1.00	0.20	1.00	0.60	0.60
4,000	1.00	1.00	0.00	0.50	0.50	1.00	0.50	0.50	0.50	0.50	1.00	0.00	1.00	0.50	0.50
4,500	1.00	1.00	0.00	0.50	0.50	1.00	0.50	0.50	0.50	0.50	1.00	0.00	1.00	0.50	0.50
5,000	1.00	1.00	0.00	0.50	0.50	1.00	0.55	0.50	0.52	0.52	1.00	0.10	1.00	0.55	0.55

TABLE 5 | Experimental results using plant dataset.

	ANN Algorithm					DT Algorithm					Proposed H-DNN Algorithm				
IR	**Acc**	**SP**	**SE**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**
1	0.51	1.00	0.96	0.99	0.98	0.73	0.97	0.96	0.97	0.97	0.95	0.94	0.97	0.95	0.95
100	0.99	1.00	0.55	0.76	0.49	0.99	0.88	0.77	0.82	0.68	1.00	0.77	1.00	0.87	0.87
500	1.00	1.00	0.36	0.50	0.50	1.00	0.79	0.68	0.63	0.62	1.00	0.58	1.00	0.76	0.74
1,000	1.00	1.00	0.45	0.50	0.50	1.00	0.77	0.73	0.61	0.59	1.00	0.53	1.00	0.71	0.67
1,500	1.00	1.00	0.54	0.50	0.50	1.00	0.67	0.77	0.58	0.57	1.00	0.33	1.00	0.65	0.64
2,000	1.00	1.00	0.50	0.50	0.50	1.00	0.68	0.75	0.60	0.61	1.00	0.36	1.00	0.70	0.73
2,500	1.00	1.00	0.33	0.50	0.50	1.00	0.64	0.67	0.57	0.57	1.00	0.27	1.00	0.64	0.64
3,000	1.00	1.00	0.11	0.50	0.50	1.00	0.71	0.56	0.57	0.55	1.00	0.43	1.00	0.64	0.61
3,500	1.00	1.00	0.00	0.50	0.50	1.00	0.80	0.50	0.61	0.58	1.00	0.60	1.00	0.71	0.67
4,000	1.00	1.00	0.00	0.50	0.50	1.00	0.80	0.50	0.62	0.61	1.00	0.60	1.00	0.75	0.71
4,500	1.00	1.00	0.00	0.50	0.50	1.00	0.58	0.50	0.54	0.54	1.00	0.17	1.00	0.58	0.58
5,000	1.00	1.00	0.00	0.50	0.50	1.00	0.61	0.50	0.56	0.56	1.00	0.22	1.00	0.61	0.61

TABLE 6 | ANN and DT results using human dataset.

	ANN Algorithm					DT Algorithm					Proposed H-DNN Algorithm				
IR	**Acc**	**SP**	**SE**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**	**Acc**	**SE**	**SE**	**F1**	**Prec**
1	0.50	1.00	0.99	0.86	0.85	0.72	0.97	0.97	0.90	0.90	0.94	0.94	0.95	0.94	0.94
100	0.99	1.00	0.79	0.50	0.50	0.99	0.85	0.89	0.67	0.66	0.99	0.69	1.00	0.83	0.83
500	1.00	1.00	0.56	0.50	0.50	1.00	0.75	0.78	0.63	0.63	1.00	0.51	1.00	0.76	0.76
1,000	1.00	1.00	0.46	0.50	0.50	1.00	0.67	0.73	0.61	0.65	1.00	0.34	1.00	0.72	0.81
1,500	1.00	1.00	0.41	0.50	0.50	1.00	0.76	0.71	0.62	0.61	1.00	0.53	1.00	0.74	0.72
2,000	1.00	1.00	0.45	0.50	0.50	1.00	0.79	0.72	0.63	0.62	1.00	0.57	1.00	0.77	0.75
2,500	1.00	1.00	0.24	0.50	0.50	1.00	0.75	0.62	0.59	0.57	1.00	0.50	1.00	0.69	0.65
3,000	1.00	1.00	0.13	0.50	0.50	1.00	0.80	0.56	0.58	0.56	1.00	0.60	1.00	0.67	0.62
3,500	1.00	1.00	0.00	0.50	0.50	1.00	0.60	0.50	0.55	0.55	1.00	0.20	1.00	0.60	0.60
4,000	1.00	1.00	0.00	0.50	0.50	1.00	0.80	0.50	0.65	0.65	1.00	0.60	1.00	0.80	0.80
4,500	1.00	1.00	0.00	0.50	0.50	1.00	1.00	0.50	0.56	0.54	1.00	1.00	1.00	0.62	0.57
5,000	1.00	1.00	0.00	0.50	0.50	1.00	0.92	0.50	0.63	0.60	1.00	0.83	1.00	0.76	0.69

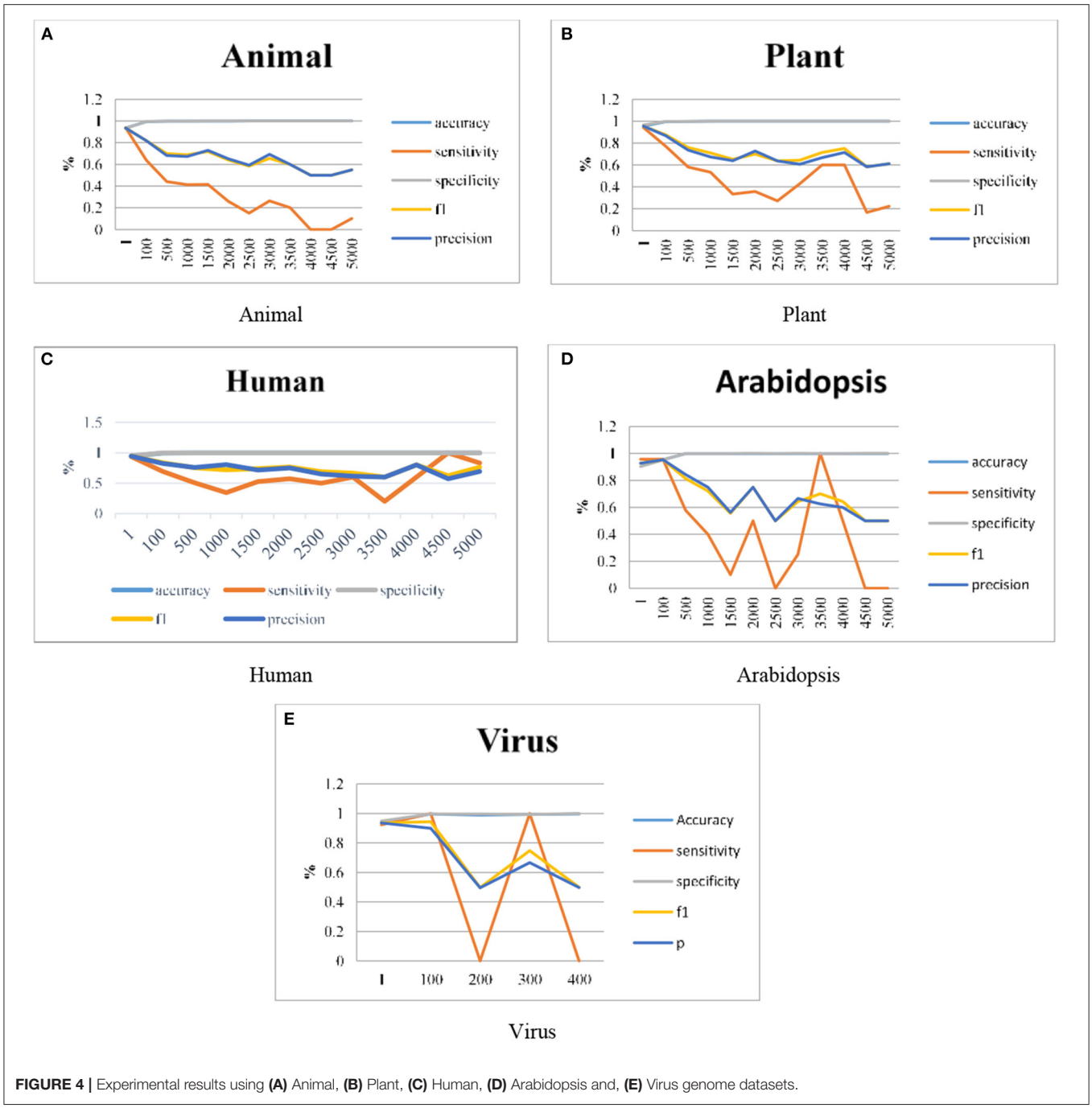


FIGURE 4 | Experimental results using **(A)** Animal, **(B)** Plant, **(C)** Human, **(D)** Arabidopsis and, **(E)** Virus genome datasets.

the ANN we get a better result for almost all the IRs up to 1:5000 as compared to the performance gained by DT and ANN separately.

CONCLUSION

The production of imbalanced data in large quantities in Bioinformatics provides a great scope for work in the AI world. These kinds of problems require modern solutions for which many questions will arise that have to be justified and have to satisfy all the complex conditions too. One such problem is the classification of these miRNA sequences which are very few in genome data and are similar to other RNA structures. The miRNA seems to help detect and diagnose cancer disease. Hence classifying this miRNA is a big challenge that has to be faced. So, a novel solution is required with minimal complexity that provides better results in terms of the TPR, time, and accurate classification of miRNA from the imbalanced genome dataset. In this paper, we have proposed a Hybrid Neural Network model in which we have integrated Deep ANN and Deep Decision Tree

TABLE 7 | ANN and DT results using arabidopsis dataset.

	ANN Algorithm					DT Algorithm					Proposed H-DNN Algorithm				
IR	**Acc**	**SP**	**SE**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**	**Acc**	**SE**	**SE**	**F1**	**Prec**
1	0.54	1.00	0.85	0.35	0.27	0.74	0.98	0.87	0.64	0.60	0.93	0.96	0.90	0.93	0.93
100	0.54	1.00	0.75	0.35	0.27	0.75	0.98	0.85	0.65	0.61	0.95	0.96	0.95	0.95	0.95
500	1.00	1.00	0.12	0.50	0.50	1.00	0.79	0.56	0.66	0.67	1.00	0.58	1.00	0.81	0.84
1,000	1.00	1.00	0.25	0.50	0.50	1.00	0.70	0.63	0.61	0.62	1.00	0.40	1.00	0.72	0.75
1,500	1.00	1.00	0.55	0.50	0.50	1.00	0.55	0.77	0.53	0.53	1.00	0.10	1.00	0.56	0.56
2,000	1.00	1.00	0.24	0.50	0.50	1.00	0.75	0.62	0.62	0.62	1.00	0.50	1.00	0.75	0.75
2,500	1.00	1.00	0.00	0.50	0.50	1.00	0.50	0.50	0.50	0.50	1.00	0.00	1.00	0.50	0.50
3,000	1.00	1.00	0.00	0.50	0.50	1.00	0.63	0.50	0.57	0.58	1.00	0.25	1.00	0.64	0.67
3,500	1.00	1.00	0.00	0.50	0.50	1.00	1.00	0.50	0.60	0.56	1.00	1.00	1.00	0.70	0.63
4,000	1.00	1.00	0.00	0.50	0.50	1.00	0.75	0.50	0.57	0.55	1.00	0.50	1.00	0.64	0.60
4,500	1.00	1.00	0.00	0.50	0.50	1.00	0.50	0.50	0.50	0.50	1.00	0.00	1.00	0.50	0.50
5,000	1.00	1.00	0.00	0.50	0.50	1.00	0.50	0.50	0.50	0.50	1.00	0.00	1.00	0.50	0.50

TABLE 8 | ANN and DT results using virus dataset.

	ANN Algorithm					DT Algorithm					Proposed H-DNN Algorithm				
IR	**Acc**	**SP**	**SE**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**	**Acc**	**SE**	**SP**	**F1**	**Prec**
1	0.52	0.48	0.99	0.52	0.52	0.86	0.94	0.92	0.86	0.92	0.94	0.92	0.95	0.94	0.94
100	0.99	1.00	0.32	0.50	0.49	0.99	1.00	0.92	0.90	0.93	1.00	1.00	1.00	0.94	0.90
200	0.99	1.00	0.16	0.50	0.50	0.99	1.00	0.96	0.48	0.44	0.99	0.00	1.00	0.50	0.50
300	1.00	1.00	0.00	0.50	0.50	0.99	0.00	0.98	0.76	0.65	0.99	1.00	0.99	0.75	0.67
400	1.00	1.00	0.00	0.50	0.50	1.00	0.00	1.00	0.41	0.49	1.00	0.00	1.00	0.50	0.50

Classifiers. The Deep ANN gives better results when compared with the existing result and good TPR with 2 epochs and up to IR 1:2000. Furthermore, when the model needs to be trained with IR up to 1:5000, Deep Decision Tree Classifier gives better results with good TPR than ANN. When we integrate both Deep ANN and Deep DT we get better performance as compared to both taken individually. In our future work, we would like to further extend by increasing the IR and reducing the time of convergence. We would also like to check the performance of the proposed model on other large genome datasets. The main aim is that the model should not follow any sort of sampling techniques and predict the miRNAs with good TPR.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

ER, SR, and RE contributed in problem formulation, implementation and results. KK, SP, SV, VV, and AM contributed in methodology identification and techniques to project results. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Peng Y, Croce CM. The role of microRNAs in human cancer. *Signal Transduct Target The*. (2016) 1:2857–67. doi: 10.1038/sigtrans.2015.4
2. Bertoli G, Cava C, Castiglioni I. MicroRNAs: new biomarkers for diagnosis, prognosis, therapy prediction, and therapeutic tools for breast cancer. *Theranostics*. (2015) 10:1122–43. doi: 10.7150/thno.11543
3. Li L, Xu J, Yang D, Tan X, Wang, H. Computational approaches for microRNA studies: A review. *Mamm Genome*. (2010) 21:1–12. doi: 10.1007/s00335-009-9241-2
4. Allmer J, Yousef M. Computational methods for ab initio detection of micrornas. *Front Genet*. (2012) 3:209–12. doi: 10.3389/fgene.2012.00209
5. Bugnon LA, Yones C, Milone DH, Stegmayer G. Deep neural architectures for highly imbalanced data in bioinformatics. *IEEE Trans Neural Netw Learn Syst*. (2019) 31:2857–67. doi: 10.1109/TNNLS.2019.2914471
6. Gudy's, Szczeániak MW, Sikora M, Makałowska I. HuntMi: An efficient and taxon-specific approach in pre-miRNA identification. *BMC Bioinf*. (2013) 14:83. doi: 10.1186/1471-2105-14-83
7. Zheng X, Xu S, Zhang Y, Huang X. Nucleotide-level convolutional neural networks for pre-mirna classification. *Sci Rep*. (2019) 9:1–6. doi: 10.1038/s41598-018-36946-4
8. Wang S, Yao, X. Multiclass imbalance problems: Analysis and potential solutions. *IEEE Trans Syst Man Cybern B Cybern*. (2012) 42:1119–30. doi: 10.1109/TSMCB.2012.2187280

9. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. (2009) 21:1263–84. doi: 10.1109/TKDE.2008.239
10. Alazab M, Khan S, Krishnan SSR, Pham QV, Reddy MPK. A multidirectional LSTM model for predicting the stability of a smart grid. *IEEE Access*. (2020) 8:85454–63. doi: 10.1109/ACCESS.2020.2991067
11. Lin M, Tang K, Yao X. Dynamic sampling approach to training neural networks for multiclass imbalance classification. *IEEE Trans Neural Netw Learn Syst*. (2013) 24:647–60. doi: 10.1109/TNNLS.2012.2228231
12. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst Appl*. (2017) 73:220–39. doi: 10.1016/j.eswa.2016.12.035
13. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. *Nature Rev Cancer*. (2015) 15:321–33. doi: 10.1038/nrc3932
14. Saçar MD, Allmer J. Machine learning methods for microRNA gene prediction. *Methods Mol Bio*. (2014) 1107:177–87. doi: 10.1007/978-1-62703-748-8_10
15. Shukla V, Varghese VK, Kabekkodu SP, Mallya S, Satyamoorthy K. A compilation of Web-based research tools for miRNA analysis. *Brief Funct Genomics*. (2017) 16:249–73. doi: 10.1093/bfgp/elw042
16. Xue C, Li F, He T, Liu, G-P, Li Y, Zhang X. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinf*. (2005) 6:310. doi: 10.1186/1471-2105-6-310
17. Helvik SA, Snøve O, Sætrom P. Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics*. (2007) 23:142–9. doi: 10.1093/bioinformatics/btl570
18. Sheng Y, Engström PG, Lenhard B. Mammalian MicroRNA prediction through a support vector machine model of sequence and structure. *PLoS ONE*. (2007) 2:e946. doi: 10.1371/journal.pone.0000946
19. Hertel J, Stadler PF. Hairpins in a Haystack: Recognizing microRNA precursors in comparative genomics data. *Bioinformatics*. (2006) 22:e197–202. doi: 10.1093/bioinformatics/btl257
20. Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, et al. Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinf*. (2005) 6:267. doi: 10.1186/1471-2105-6-267
21. Huang T-H, Fan B, Rothschild M, Hu ZL, Li K, et al. MiRFinder: An improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinf*. (2007) 8:341. doi: 10.1186/1471-2105-8-341
22. Ding J, Zhou S, Guan J. MiRenSVM: towards better prediction of microRNA precursors using an ensemble SVM classifier with multiloop features. *BMC Bioinf*. 11:S11. doi: 10.1186/1471-2105-11-S11-S11
23. Batuwita R, Palade V. microPred: Effective classification of premiRNAs for human miRNA gene prediction. *Bioinformatics*. (2009) 25:989–95. doi: 10.1093/bioinformatics/btp107
24. Kleftogiannis D, Theofilatos K, Likothanassis S, Mavroudi, S. YamiPred: a novel evolutionary method for predicting pre-miRNAs and selecting relevant features. *IEEE/ACM Trans Comput Biol Bioinf*. (2015) 12:1183–92. doi: 10.1109/TCBB.2014.2388227
25. Pandya S, Ghayvat H. Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence. *Adv Eng Inform*. (2021) 47:1012. doi: 10.1016/j.aei.2020.101238
26. Stegmayer G, Yones C, Kamenetzky L, Milone DH. High class-imbalance in pre-miRNA prediction: A novel approach based on deepSOM. *IEEE/ACM Trans Comput Biol Bioinf*. (2017) 14:1316–26. doi: 10.1109/TCBB.2016.2576459
27. Fischer A, Igel C. An introduction to restricted Boltzmann machines. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (Lecture Notes in Computer Science)*. Berlin, Germany: Springer (2012).
28. Ayachit G, Shaikh I, Pandya H, Das J. Salient features data and algorithms for microRNA screening from plants: a review on the gains and pitfalls of machine learning techniques. *Curr Bioinform*. (2020) 15:1091–103. doi: 10.2174/1574893615999200601121756
29. Reddy GT, Reddy MPK, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G. Analysis of dimensionality reduction techniques on big data. *IEEE Access*. (2020) 8:54776–88. doi: 10.1109/ACCESS.2020.2980942
30. Koumakis L. Deep learning models in genomics; are we there yet? *Comput Struct Biotechnol J*. (2020) 18:1466–73. doi: 10.1016/j.csbj.2020.06.017
31. Shi H, Wang D, Wu P, Cao Y, Chen, Y. Deep Multiple Kernel Learning for Prediction of MicroRNA Precursors. *Sci Program*. (2021) 2021. doi: 10.1155/2021/9969282
32. Peace RJ, Biggar KK, Storey KB, Green JR. A framework for improving microrna prediction in non-human genomes. *Nucleic Acids Res*. (2015) 43:e138. doi: 10.1093/nar/gkv698
33. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, et al. CP-BDHCA: blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform*. (2021) 6:1–1. doi: 10.1109/JBHI.2021.3097237
34. Liu B, Fang L, Chen J, Liu F, Wang X. miRNA-dis: MicroRNA precursor identification based on distance structure status pairs. *Molecular BioSyst*. (2015) 11:1194–204. doi: 10.1039/C5MB00050E
35. Wei L, Liao M, Gao Y, Ji R, He Z, Zou Q. Improved and promising identification of human microRNAs by incorporating a highquality negative set. *IEEE/ACM Trans Comput Biol Bioinf*. (2014) 11:192–201. doi: 10.1109/TCBB.2013.146
36. Ghayvat H, Awais M, Pandya S, Ren H, Akbarzadeh S, Chandra Mukhopadhyay S, et al. Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. Sensors. (2019) 19:766. doi: 10.3390/s19040766
37. Mahmud, Kaiser MS, Hussain A, Vassanelli S. Applications of deep learning and reinforcement learning to biological data. *IEEE Trans Neural Netw Learn Syst*. (2018) 29:2063–79.
38. Si Z, Yu H, Ma Z. Learning deep features for DNA methylation data analysis. *IEEE Access*. (2016) 4:2732–7. doi: 10.1109/ACCESS.2016.2576598
39. Stegmayer G, Gerard M, Milone D. Data mining over biological datasets: An integrated approach based on computational intelligence. *IEEE Comput Intell Mag*. (2012) 7:22–34. doi: 10.1109/MCI.2012.2215122
40. LeRoux N, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput*. 20:1631–1649. doi: 10.1162/neco.2008.04-07-510
41. Thomas J, Thomas S, Sael, L. DP-miRNA: an improved prediction of precursor microRNA using deep learning model. *In Proc IEEE Int Conf Big Data Smart Comput*. (2017) 1:96–9. doi: 10.1109/BIGCOMP.2017.7881722
42. Lertampaiporn S, Thammarongtham C, Nukoolkit C, Kaewkamnerdpong B, Ruengjitchatchawalya M. Heterogeneous ensemble approach with discriminative features and modified-SMOTEbagging for pre-miRNA classification. *Nucleic Acids Res*. (2013) 41:e21. doi: 10.1093/nar/gks878
43. Bugnon LA, Yones C, Milone DH, Stegmayer G. Genome-wide discovery of pre-miRNAs: comparison of recent approaches based on machine learning. *Brief Bioinformatics*. (2021) 22:bbaa184. doi: 10.1093/bib/bbaa184
44. Tasdelen A, Sen B. A hybrid CNN-LSTM model for pre-miRNA classification. *Sci Rep*. (2021) 11:1–9. doi: 10.1038/s41598-021-93656-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 R., Jain, Kotecha, Pandya, Reddy, E., Varadarajan, Mahanti and V. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fpubh.2021.788376

Machine Learning Assisted Cervical Cancer Detection

Mavra Mehmood[1], *Muhammad Rizwan*[1], *Michal Gregus ml*[2*] *and Sidra Abbas*[3*]

[1] *Department of Computer Science, Kinnaird College for Women, Lahore, Pakistan,* [2] *Information Systems Department, Faculty of Management, Comenius University in Bratislava, Bratislava, Slovakia,* [3] *ASETS Lab, Islamabad, Pakistan*

Cervical malignant growth is the fourth most typical reason for disease demise in women around the globe. Cervical cancer growth is related to human papillomavirus (HPV) contamination. Early screening made cervical cancer a preventable disease that results in minimizing the global burden of cervical cancer. In developing countries, women do not approach sufficient screening programs because of the costly procedures to undergo examination regularly, scarce awareness, and lack of access to the medical center. In this manner, the expectation of the individual patient's risk becomes very high. There are many risk factors relevant to malignant cervical formation. This paper proposes an approach named *CervDetect* that uses machine learning algorithms to evaluate the risk elements of malignant cervical formation. *CervDetect* uses Pearson correlation between input variables as well as with the output variable to pre-process the data. *CervDetect* uses the random forest (RF) feature selection technique to select significant features. Finally, *CervDetect* uses a hybrid approach by combining RF and shallow neural networks to detect Cervical Cancer. Results show that *CervDetect* accurately predicts cervical cancer, outperforms the state-of-the-art studies, and achieved an accuracy of 93.6%, mean squared error (MSE) error of 0.07111, false-positive rate (FPR) of 6.4%, and false-negative rate (FNR) of 100%.

Keywords: cervical cancer, medical data, gynecological diseases, artificial intelligence, feature engineering, classification

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Weizheng Wang,
City University of Hong Kong, Hong Kong SAR, China
Giridhar Bojja,
Dakota State University, United States

***Correspondence:**
Sidra Abbas
sidra.abbas708@gmail.com
Michal Gregus ml
michal.gregus.ml@fm.uniba.sk

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *02 October 2021*
Accepted: *16 November 2021*
Published: *23 December 2021*

Citation:
Mehmood M, Rizwan M, Gregus ml M and Abbas S (2021) Machine Learning Assisted Cervical Cancer Detection. Front. Public Health 9:788376. doi: 10.3389/fpubh.2021.788376

1. INTRODUCTION

Machine learning (ML) and deep learning (DL) are being utilized for brain tumor detection, cervical cancer detection, breast cancer detection, COVID detection, physical activity recognition, thermal sensation detection, and cognitive health assessment of dementia individuals (1–3, 3–10). Advancements in Health Care Industry makes it more effective than traditional diagnosing techniques (11–14). According to medical reports published by Global cancer statistics (15) every year, 493,000 cervical malignancy patients have been added, among which 15% are female malignancy patients. This disease is mainly found in developing countries (16) with an 83% death ratio. Prominent in African countries (17) e.g., Uganda, which ranks fourteenth among the highest incidence of cervical malignancies with 65% of confirmed cases (18).

Human papillomavirus (HPV) contamination is mainly caused by cervical malignancy. HPV contamination is sexually transmitted. Sexual behavior associated with age at first sexual contact and the sexual activity of the accomplice is linked to the increased risk of HPV acquisition. Throughout this case, the cervical malignancy is profoundly preventable through accessible screening and detection than other different sorts of malignancy, and it is critical to actualizing

risk expectations. The malignant cervical formation is a malignant tumor. Cervical tissue cells expand and replicate abnormally without regulated cell division and death cause. If the tumor becomes malignant, the cell travels to other areas of the body, such that certain sections become subsequently infected and, in most serious situations, may be avoided through early identification (19). Deaths due to cervical malignancies can be reduced if effective screening strategies are implemented (20). With the fast advancement of present-day clinical innovation and computer technological innovation, various screening and diagnostic strategies depend on computer-aided (CAD) architectures.

A procedure of retrieving applicable information from information sources is known as data mining (21, 22). Real-world information comprises grimy information, for example, inaccurate and incomplete. Along these lines, cleaning and changing crude information to permit a reliable analytic delivery can accurately reflect the outcome (23, 24). It is implemented on the dataset. The cervical cancer dataset acquired for examination contains redundancy, missing values, and noise. Mining methods are considered one of the greatest challenges and significant fields of study in medicine due to the growing importance of health issues (25). The data mining framework can advance the cervical cancer screening process with the help of the knowledge it extracts (26). Across the medical sector, these techniques are helpful not just in finding similarities and correlations between such symptoms but also in predicting diseases. By applying several data mining techniques, continuous research and medical treatment can be immediately recommended, resulting in life-saving, especially for cervical cancer. The first step is pre-processing 80% of information, which plays a vital role as a round of all data mining operations (27).

The random forest (RF) is used to identify essential features in our research that improve the training dataset's quality. RF constructs a forest of tree classifications in which each tree grows on a bootstrap sample data, and the characteristics of every other tree node are selected from a random subset of all characteristics. The final level of the entity is measured by voting across all trees in the forest. Several significant benefits to an RF method makes it the perfect methodology for studying particular biological data in pharmacogenomics research findings. Next, it can accommodate a wide range of both qualitative and quantitative input vectors. Second, it tests the attribute's importance in estimating the type, thereby providing a benchmark for selecting features. Third, RF generates an accurate classifier for unbiased internal generalized analysis during the forest growth process. Finally, RF is relatively stable in the face of etiological variability, and a reasonably low amount of missed data (28).

This article makes the following contributions:

- *CervDetect* implements ML algorithms on the medical data to gain a deeper understanding and to evaluate the risk elements of malignant cervical formation.
- Utilizes Pearson correlation between input variables and the output variable to pre-process the data.
- *CervDetect* uses the RF feature selection technique to select significant features.
- Proposes a Hybrid approach by combining RF and shallow neural network to detect cervical cancer.
- Effectively enhance the detection rate in comparison with state-of-the-art studies.

The rest of the part is arranged as follows. Section 2 presents the literature review. Section 3 presents the dataset and preliminaries. Section 4 presents the proposed methodology. Section 5 examines the result and discussion. Finally, section 6 presents the conclusion and future work.

2. LITERATURE REVIEW

In earlier studies, various methodologies based on traditional ML approaches, including k-nearest neighbors (KNN) and K-means clustering and RF, have been utilized for cervical cancer diagnosis (29–31). In 2017, WEN WU1 and HAO developed a clinical decision support network for malignant formation, using a knowledge-based method in conjunction with a rough collection of hypotheses hereditary genetic algorithms calculations in a soft computing model (32). Parallelism, self-learning, and sensitivity to internal failure are artificial neural network (ANN) characteristics. The suggested paradigm takes advantage of these characteristics. Although the information processing capabilities of the rough set theory, and the robust, parallel, and vigorous search are characteristics of the GA's. It still faces high computation problems.

Ashok, Dr. P. Aruna (33) discussed various methods of ML, including support vector machines (SVM), gray level concurrence matrix (GLCM), KNN, convolutional neural networks (CNN), spatial fuzzy cluster algorithms, RF, C5.0, and hierarchical cluster algorithm for feature extraction, segmentation, cell classification, and evaluated in typical parameters, such as dataset volume, disadvantages, and precision. Although it is only best suited for small datasets. Chih-Jen Tseng and Chi-Jie Lu (34) have used image processing, data extraction, and ML techniques to diagnose cervical cancer. The combination of texture and shape features is extracted from each image. Optimal properties are chosen to select reciprocal information, sequential forward selection, and random section selection. Images of cervical cancer are classified using the SVM. Various selection methods are contrasted to determine the proper mechanism suitable for the diagnosis of cervical cancer. The problem faced by SVM is that there is no probabilistic explanation for classification, which leads to very rigid classifications.

In 2006, Reif et al. (35) described RF efficacy in various model genetic and proteomic datasets. RF progress fails to classify related traits based on genetic data and proteomics datasets. Experimental findings indicate that using several data sources is useful where the disease definition is uncertain, and the corresponding data basis for the phenotypic outcome is unknown. This study's findings indicate that RF is exceptional for detecting high-dimensional data vector characteristics with minimal main effects and low heritability, but the problem faced by RF chooses only one attribute at each tree split during construction, strictly epistatic. In 2016, Vidya and Nasira (36) worked on predicting the normal cervix. The cancer cervix is

evaluated using practical data mining algorithms. Geetha and Thangamani (37) addressed the imbalanced distribution of data and risk factors for cervical cancer diagnosis.

Anuraga et al. (38) attempted to explore the conditions that affect cancer patients' survival in Makassar, Indonesia. The specimens included in this research contained as many as 38 cancer sufferers. They use the RF to identify tree merger data by combining sample data training. The main problem faced by this technique is that it only achieves 50% of accuracy. The Bandyopadhyay and Nasipuri (39) focused on K-Means clustering to segmentation pre-processed images, and Herlev analysis is carried out. IOU segmentation outcomes, verified ground truth values, and shape attributes are extracted from the segmented nucleus. The identification of the nucleus on the grounds of the shape features is carried out with the RF Classifier's aid, and the contrast is rendered with the other classifiers.

In 2020, Alyafeai and Ghouti (40) build a completely integrated cervical cancer identification and cervical cancer screening pipeline from cervical images. The current pipeline comprises two deep neural network-learning models for automated cervical identification and diagnosis of cervical tumors. The first test detects the cervix area 1,000 times faster than the state-of-the-art data-driven simulations, thus obtaining a detection precision of 0.68 in terms of union intersection (IoU) estimation. Self-extracted characteristics are used in the second model to identify cervical tumors. Such features are trained using two lightweight models focused on co-evolutionary neural networks (CNNs). William et al. (41) performed to reduce the probability of mistake by automating the diagnostic process for cervical cancer from Pap-drug photos. Local adaptive histogram equalization was used for image enhancement.

Brats data (24) is used to localize brain tumors in FLAIR scans of MRI. Bilateral flipping is applied to remove noise, while the Gabor filter bank creates text-ton map images. A superpixel is generated by segmenting out an image superpixel is generated by segmenting-level features include at each superpixel histogram level of texton-map is calculated, and the first-order intensity features are calculated as well. Their major contribution is that they made low-level features significant for the localization of brain tumors at the region level approach by integrating features from texton-map images. The prediction of three classes, namely tumor, non-tumor, and background, was made by giving these extracted features later to the classifier. A cross-validation technique leave-one-out (LOOCV) for the tumor area localization is applied to outperform the existing solution and achieve a dice score of 88%.

In 2021, Wang et al. (42) worked to solve challenges regarding physical security and over-centralized server problems in wireless medical sensor networks (WMSN). Physically unclonable function (PUF) and cutting-edge blockchain technology are combined to propose a reliable and authentic protocol for WSMN. In addition, to deal with biometric information, fuzzy extractor method has been used. For the evaluation of their proposed method, two schemes are used. Their work outperforms the existing studies by achieving higher accuracy through minimizing computation and communication costs. Similarly, Xiong et al. (43) provided a design of blockchain-based ECDSA with fault-tolerant batch verification protocol for blockchain-enabled IoMT. In 2021, Khamparia et al. (1) combined a convolutional network with a variational encoder for data classification. The dimensionality of images data can be reduced by using a variational encoder with a softmax layer with the kernel size of 2x2 and 3x3. Their architecture outperforms the current ML models. Chen et al. (11) developed Cyto Brain that facilitates in subsequent clinical diagnosis, an artificial intelligence (AI) based system. CytoBrain consists of three main modules: (1) to extract only cell images in a whole slide image efficiently, cervical cell segmentation module has been designed. (2) for the cell classification, vgg 16 is used, and a classifier module is designed; Moreover, the last one is the human-aided diagnosis module which can automatically diagnose cervical cancer based on the classification results of cells on a whole slide image.

3. DATASET AND PRELIMINARIES

The dataset[1] used has been released in The University of California's database at Irvine (UCI). The collection contained existing patient history, practices, and procedures and demographic statistics for 858 instances with 32 features per scenario. The dataset contains several missed features since others are incomplete instances, which were chosen not to tackle any privacy concerns, as privacy and security are the common issues in healthcare record frameworks (44–48). **Table 1** displays the dataset characteristics and the missed value for each function.

4. PROPOSED METHODS

Figure 1 illustrates the flow of our work in a block diagram consisting of pre-processing and classification of cervical cancer. The proposed approach consists of pre-processing and classification of cancer using ML algorithms and deep learning algorithms.

Algorithm 1 shows the working of *CervDetect*. Random samples have been collected from the input cervical cancer risk factor dataset. For each collected sample, a decision tree has been made to make a prediction. Counting has been done for each final score, and the features are ranked according to their importance. Selected features are input to the shallow neural network where x_i(i=1, 2, ... n) are the number of input features, w_{21} and w_{22} are the weights and B1 is constant. y stores the features convolved with a randomly initialized filter matrix. V store results of applying a linear transformation on y. O_1 stores the output generates after applying a sigmoid function on y, and e calculates the change in error concerning weights. Infeed backward change error to the weights has been calculated with the help of the chain rule, and the weights are updated.

4.1. Data Pre-processing

The dataset regarding cervical cancer has many missed values. A missing value can imply a variety of differences. Records with

[1]https://archive.ics.uci.edu/ml/index.php

TABLE 1 | Dataset description.

Serial number	Features names in dataset	No of missing values
1	"Number of sexual partners"	26
2	"First sexual intercourse"	7
3	"Num of pregnancies"	56
4	"Smokes"	13
5	"Smokes (years)"	13
6	"Smokes (packs/year)"	13
7	"Hormonal Contraceptives"	108
10	"Hormonal Contraceptives (years)"	108
11	"IUD"	117
12	"IUD (years)"	117
13	"STDs"	105
14	"STDs (number)"	105
15	"STDs:condylomatosis"	105
16	"STDs:cervical condylomatosis"	105
17	"STDs:vaginal condylomatosis"	105
18	"STDs:syphilis"	105
20	"STDs:pelvic inflammatory disease"	105
21	"STDs:genital herpes"	105
22	"STDs:molluscum contagiosum"	105
23	"STDs:AIDS"	105
24	"STDs:HIV"	105
25	"STDs:Hepatitis B"	105
26	"STDs:HPV"	105
27	"STDs: Time since first diagnosis"	787
28	"STDs: Time since last diagnosis"	787
29	"Age"	0
30	"STDs: Number of diagnosis"	0
31	"Dx:Cancer"	0
32	"Dx:CIN"	0
33	"Dx:HPV"	0
34	"Dx"	0
35	"Hinselmann"	0
36	"Schiller"	0

Algorithm 1 *CervDetect* Algorithm

Input: Cervical cancer risk factor dataset.

Begin with the collection of random samples from a dataset.
First, this algorithm must generate a decision tree for each sample.
The prediction will come from the decision tree.
Counting will be carried out in this stage for each final score.
Ultimately, pick the most elected outcome of the prediction as to the outcome of the prediction.
Output: Rank features with according to their importance.
Input: selected features input to shallow neural network x_i(i=1, 2, 3, 4…n)
Initialization is the first step after the configuration of the neural network
Initiate all weights w_{21}, w_{22} with a random number from the usual distribution, i.e. N (0, 1).
Put all of the bias nodes B1 = B2 = 1.0.
Feedforward
for $I \in l$ **do**
  each of x input have weights w21, w22 …
  y=$w_{21}\ x_1$, $w_2 2\ x_2$, …,$w_m\ x_i$,
  v= sigmoid(y)
  $o_1 = y_{v1}\ w_{21}$
  $o_v = o_1$
  e=$\frac{1}{n}\sum_{i=1}^{2}$ (y_i-ovi)^2
end for
Feed backward
for $i \in k < l$ **do**
  Partial derivative of the e to the weight adjusted w_i^{k+1} w_{ik}-Δ /(φw_{k^i})
  weight update e_i=e_i Δ+w_i^{k+1}
end for

missed values could be included, omitted, or the mean of the variable can be replaced for missing numerical characteristics or with the most frequent value in the case of categorical features. The technique of eliminating the amount of data with missing rows of values decreased from 858 to 737 when applied. We aim to scale the number of features but not the number of input vectors used in the dataset. The following steps are included to pre-process data and make it suitable input for the classifier.

1. Nan: To see relations between variables to treat Nan. Nevertheless, according to the data, the scale of more than 100 Nan values may impact results. We fill features with a median value of less than 100 Nan values in them with median values.
2. Hormonal contraception: There are so many Nan values during the data diagnosis information. Because of this, we cannot calculate the impact of this information, and must eliminate them. Instead, using the Pearson correlation, we will decide which attribute influences hormonal contraceptives. We fill Nan values with correlated attributes according to the resultant heat map. If the patient is older than the mean sample or the amount of pregnancy is less than the mean, the patient may take hormonal contraceptives. Nan values in HC (years) fill with median values using HC attributes.
3. IUD: Using the Pearson correlation, we can decide which attribute is the "IUD" influencer. Age and amount of maternity factors have an impact on the IUD feature. This indicates that 80% of patients taking IUD are older than the average age. 70% of patients who do not take IUD have fewer births than the mean amount of pregnancies. We can fill in the remaining Nan IUD feature values. With the IUD (years) rule, we can fill the -1 values with the IUD rule. If the patient takes IUD, then the UID (years) will be non-zero, so we need to adjust it to mean values.
4. STDs: condylomatosis and STDs: vulvo-perineal condylomatosis has the "STD" effect. We never consider "STD (number)" and "STDs: number of diagnoses" as they are the same attributes as "STDs." Based on our analysis of STDs,

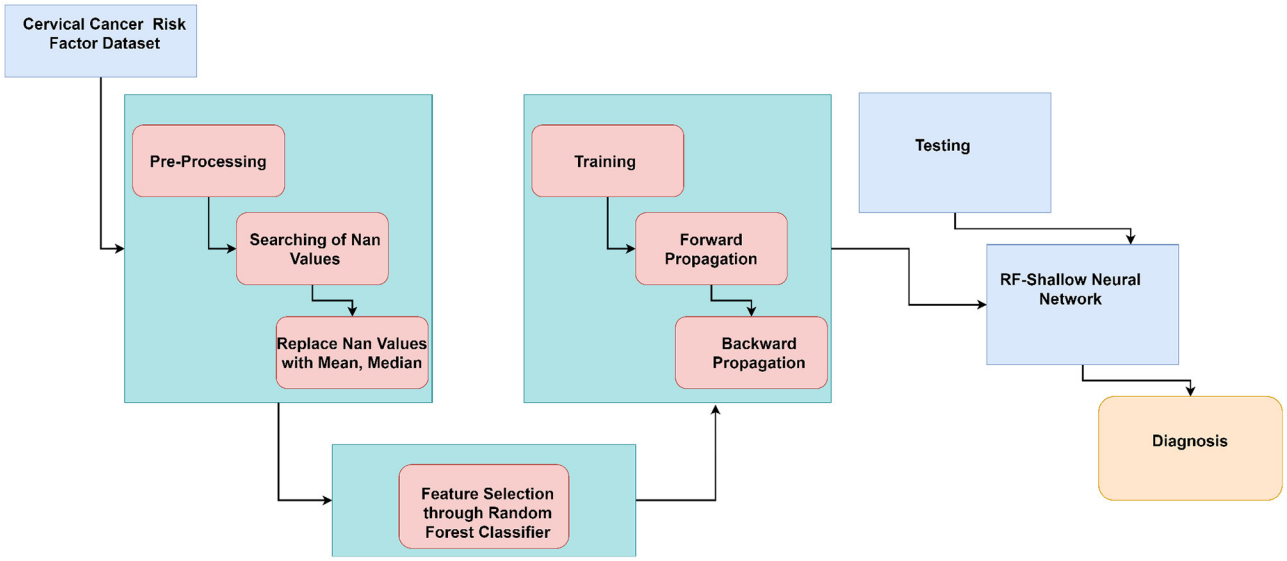


FIGURE 1 | Block diagram of proposed work flow.

we can conveniently fill the Nan values of 1 or zero, and if the individual has either of the STDs, then, the STDs should be 1, the other one will be 0.

5. STD diseases: According to information, 73% of non-smoking patients have STD. We do not have Nan values 75% of patients who do not take IUD are even STDs. Even STDs (number) are the same attribute as STDs. The median case is not helpful because we use mean values for replacement. As per the heat map and our domain awareness, all STD diseases rely on the STDs function and STD (number). We fill Nan values using median for specific values since all STDs rely on certain STDs, and we cannot be assured of a person's disease.
6. AIDS: The correlation function does not give us any hints. However, we recognize that AIDS is also a condition of STDs. So, we fill in the Nan values.
7. STDs-Hepatitis B: This aspect is the result of the STDs-HIV factor. Here we have one patient with a disease, and this importance is negligible compared to the population. The value of Nan is filled with 0.
8. STDs-HPV: The positive values of HPV are not enough. Zero values cannot be determined by using significant attributes, and we filled the value of Nan with 0.
9. STDs-time after the first diagnosis and STDs-time after the last diagnosis: When the patient's STDs are negative, the first and last diagnoses cannot be positive. This assumption fills the Nan values in these attributes. **Figure 2** illustrates the Pearson correlation between attributes in the cervical cancer risk factor dataset after handling missing values.

4.2. Random Forest

The RF model utilizes a sequence of decision trees, where each tree in the forest has been trained using a bootstrap sample of data items, so each tree splitting attribute is picked from a random subset of features. The categorization of patterns shall be based on the collective vote of all trees in the forest. Attribute importance is measured as the decreases in the node's impurity, weighted by the likelihood of approaching the node. The likelihood of the node can be determined by the number of instances that enter the node, separated by the total number of instances. The greater the value, the more important the characteristic is. In Scikit-learn Decision tree, approximates the value of the nodes using Gini significance, considering just two child nodes (binary tree):

$$nodei_{j=} Weight_j C_j - W_{l(j)} C_{l(j)} - W_{r(j)} C_{r(j)} \quad (1)$$

In Equation (1), $nodei_j$= Importance of j node; $Weight_j$= Weighted samples that reach at node j; C_j = Impurity of jth node; l(j)= From left split child node on jth node; r(j)= From right split child node on jth node. Each feature importance calculated at decision tree is given by Equation (2):

$$featureimportancei_i = \frac{\sum j\, n_{i(j)}}{\sum kn_{ik}} \quad (2)$$

Where, $\sum$j= splitting of a feature I at node j; $n_{i,j}$= node j importance; $\sum$k= all nodes. In Equation (3), normalizing these values between 0 and 1 is done by taking the sum of all feature's importance values and then dividing them.

$$normalizedfi_i = \frac{fi_{i(feature\ importance)}}{\sum j\, fi_j(all\ features\ )} \quad (3)$$

In Equation (4) at the RF stage, the final feature of importance is the sum, including all trees. Estimating an attribute's significance at each tree is calculated and divided by the overall tree species.

$$Random\ forestfi_i \frac{\sum j\ normalizedfi_{ij}}{T} \quad (4)$$

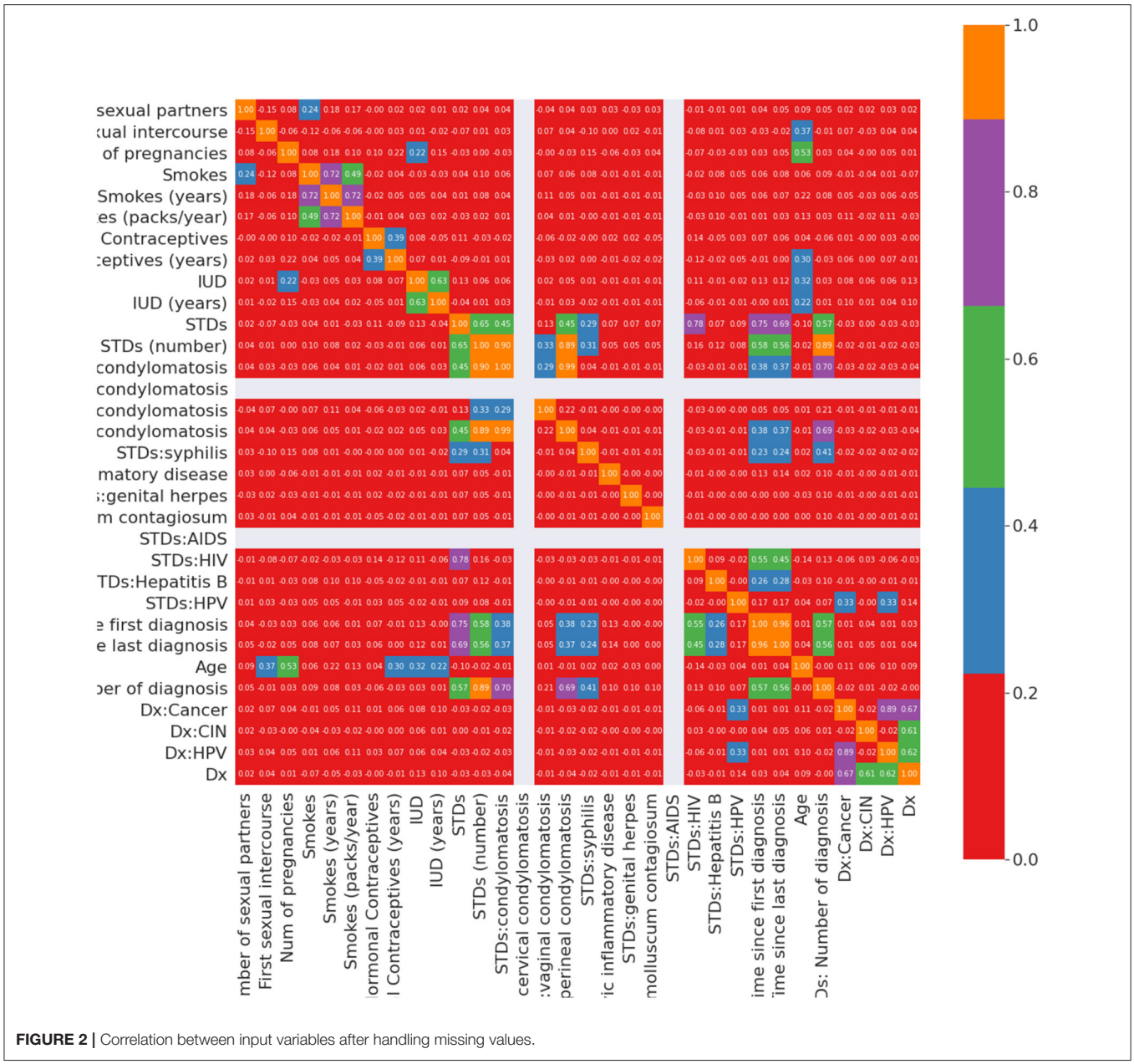


FIGURE 2 | Correlation between input variables after handling missing values.

Normalized $fi_{i,j}$ is the normalized feature importance for I feature at jth tree, and T is the total number of trees in the forest. **Figure 3** illustrates the feature importance graph that improves the quality of the training dataset.

4.3. Artificial Neural Networks

Artificial neural network models are influenced by the human brain structure, which interconnects many biological neurons that are important for maintaining coherent communication. The ANN architecture is computer-based and consists of a variety of simple parallel processing units (49). It is a common statistical technique that can analyze exact relationships between variables. There are several parallel layers in the ANN models, and each layer consists of several neurons. Input, hidden, and output layers are three different types of layers. No calculations are conducted on the input layer. Only feature variables are input. The hidden layer is the key part of the ANN and includes different neurons commonly detected by checking and error. The linear and non-linear functions for the hidden layer neurons are required. The first stage is where the hidden layer neurons receive the input variables compounded by the corresponding correlations (weights), and the second cycle is followed by a non-linear induction system, usually a sigmoid. A fully connected ANN consists of neurons. The neurons are split into layers composed of one input layer, one output layer, and numerous hidden layers with the contribution of each layer in the next

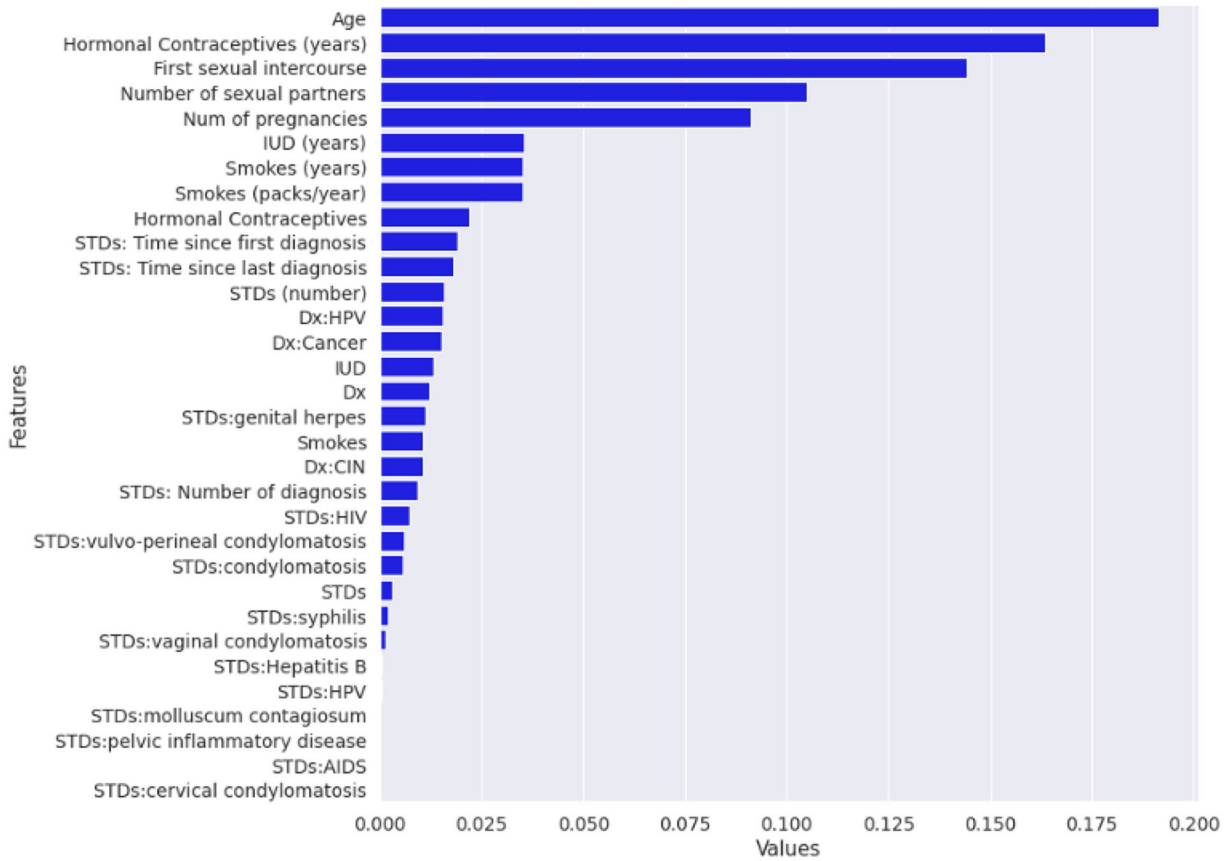


FIGURE 3 | Graph plotted after applying random forest (RF) algorithm for feature importance.

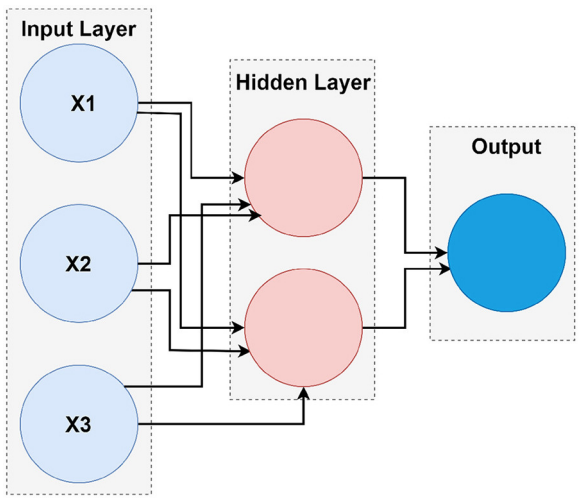


FIGURE 4 | The architecture of shallow neural network.

layer as an input in our research, and we concentrate on neural networks with only one layer of hidden data, hidden neurons, and a single output.

4.4. Shallow Neural Network Architecture

Different types of neural networks, both shallow and deep, have been built. The terms "shallow" and "deep" refer to the neural network with a minimum number of layers typically known to have a common hidden layer. Deep neural nets are used by neural networks that contain several deep layers. These types of networks execute various tasks, and the fundamental framework of shallow networks enables them to do so. Almost all shallow networks had an input layer, a single hidden layer, and an output layer. The number of hidden nodes in the layer is the only different hyper-parameter. Networks use online learning instead of batch learning, which uses simple backpropagation and gradient descent. Besides, a scaled conjugate gradient descent (SCG) backpropagation algorithm was used in shallow networks. **Figure 4** is the fundamental model for shallow neural networks.

The neurons are the atomic component of the neural network. For input data, the output is defined as well as the output is passed to the next layer as an input vector. A neuron may be thought of as a fusion of different segments:

1. The first section calculates the output Y, using input parameters and weights.
2. The second section executes the activation on Y to give the final performance of neuron A.

The hidden layer consists of several neurons, which each executes below mentioned 5 and 6 equations. Those two neurons found in the hidden layer of the external neural network are evaluated as follows:

$$Y_1^{[1]} = w_1^{[1]}Tz + b_1^{[1]}, a_1^{[1]} = \sigma\left(y_1^{[1]}\right) \quad (5)$$

$$Y_2^{[1]} = w_2^{[1]}Tz + b_2^{[1]}, a_2^{[1]} = \sigma\left(y_2^{[1]}\right) \quad (6)$$

Where the superscript number [i] is the number of the layer, and the subscription number j is the number of the neurons in a particular layer. Y is an input vector made up of three components.

W_i represents the weights connected to each input variable, and bi represents the bias factor connected to each line, hidden line, and output layer. Z[i]j is just the intermediate output associated with the j neurons found in the ith layer. A[i]j is the final output associated with the neuron j in the ith layer.

TABLE 2 | Confusion matrix of *CervDetect*.

Predicted class /Actual class	Biopsy	Normal
Biopsy	TP	FN
Normal	FP	TN

TABLE 3 | Selected optimal features.

Serial no	Important features according to their rank
1	"Age"
2	"Hormonal Contraceptives (years)"
3	"First sexual intercourse"
4	"Num of pregnancies"
5	"IUD (years)"
6	"Smokes (years)"
7	"Smokes (packs/year)"
8	"Hormonal Contraceptives"
9	"STDs: Time since first diagnosis"
10	"STDs: Time since last diagnosis"

Sigma is a function of activation that squashes the input value into the range of (0,1). Mathematically, this is described as in Equation (7):

$$\sigma = \frac{1}{1 + e^{-yi}} \quad (7)$$

Equation (8) represents all Z intermediate outputs in a single multiplication matrix.

$$Z^1 = X^{[1]T}X + b^1 \quad (8)$$

$$A_1^{[1]} = \sigma(Z^1) \quad (9)$$

The above Equation (9) represents all activation A in a single multiplication matrix.

To calculate the output for an input vector y, the following steps are performed as mentioned in below (10), (11), (12), and (13) equations. These steps can also be called feed-forward propagation.

$$Z^1 = W^{[1]T}Y + b^1 \quad (10)$$

$$A_1^{[1]} = \sigma(Z^1) \quad (11)$$

$$Z^2 = W^{[2]T}A^1 + b^2 \quad (12)$$

$$Z^{final} = A^{[2]} = \sigma(Z^2) \quad (13)$$

Where the Equation (10) measures the Z (18) intermediate output of the first hidden layer. The Equation (11) is used to measure the final production A (18) of the first hidden layer. The Equation (12) calculates the Z intermediate value of a processing layer Z. The Equation (13) calculates the end product A of the

TABLE 4 | Accuracy table.

serial no	Performance measures	Performance%
1	Accuracy	93.6%
2	TPR	100%
3	FPR	100%

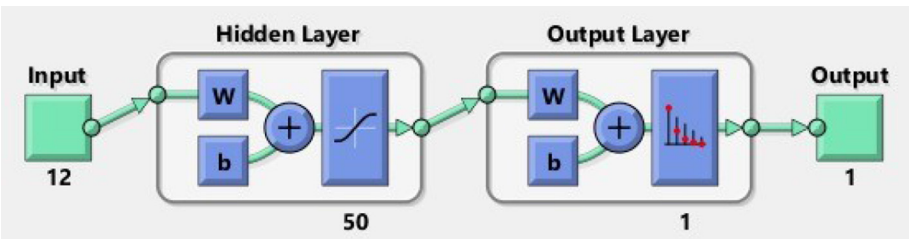


FIGURE 5 | Shallow neural network.

output layer, which is now the result of the whole neural network. When the results are obtained from each neuron's hidden layer, they are transferred to the next layer, where each neuron in the output layer finalizes the values.

$$Y_1^{[2]} = w_1^{[2]} out\,(y_1) + w_2^{[2]} out\,(y_2) + .. + w_{14}^{[2]} out(y_{14}) + b_2^{[2]} \quad (14)$$

$$a_1^{[2]} = \sigma\left(Y_1^{[2]}\right) \quad (15)$$

$$Y_{15}^{[2]} = w_{30}^{[2]} out\,(y_1) + w_{31}^{[2]} out\,(y_2) + .. + w_{43}^{[2]} out(y_{15}) + b_2^{[2]} \quad (16)$$

$$a_{50}^{[2]} = \sigma\left(Y_{50}^{[2]}\right) \quad (17)$$

Equations (14– 17) are used to determined the error after the results are obtained for each neuron in the output layer. When this is the critical error, it will avoid back-propagating to change the previous weights to get the minimal error in the feed-forward process.

5. RESULTS AND DISCUSSION

In our work, the most significant performance measure for cervical cancer diagnosis accuracy is used to calculate the RF-shallow neural network's performance. The true positive (TP) is identical to those rejected, representing the number of cancer patients marked as Biopsy. False-positive (FP) is the inverse to deny wrongly and represents normal patients as cancer patients. The true negative (TN) is equal to those correctly identified, representing the number of normal patients identified as normal. False-negative (FN) is equal to any incorrectly identified, representing the number of cancer patients identified as normal patients. **Table 2** shows the confusion matrix.

Accuracy: the proportion of the number of patient records correctly categorized against the overall number of patient records in the dataset as shown in Equation (18).

$$AC = \frac{TP + TN}{Tp + TN + FP + FN} \quad (18)$$

True positive ratio (TPR): it is identical to detection rate (DR). TPR indicates the proportion of the number of patient records

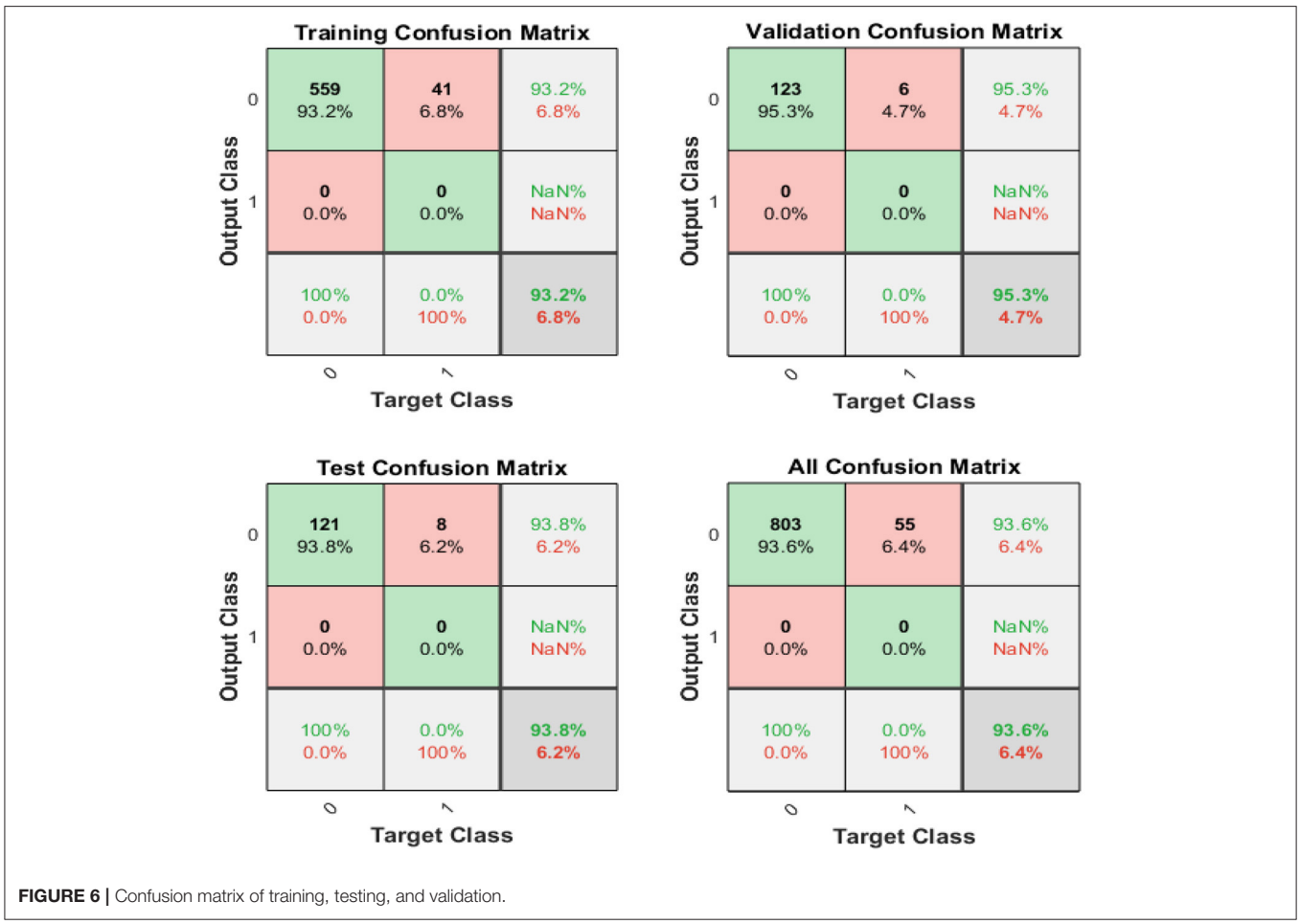


FIGURE 6 | Confusion matrix of training, testing, and validation.

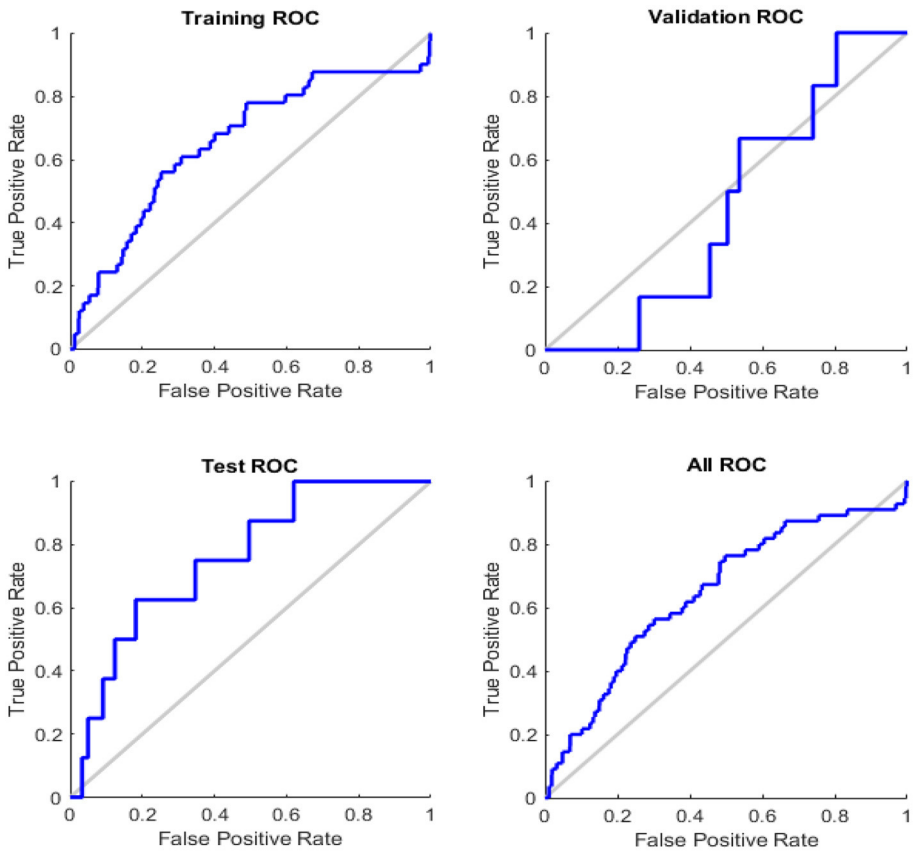


FIGURE 7 | Roc curve of testing, training, and validation.

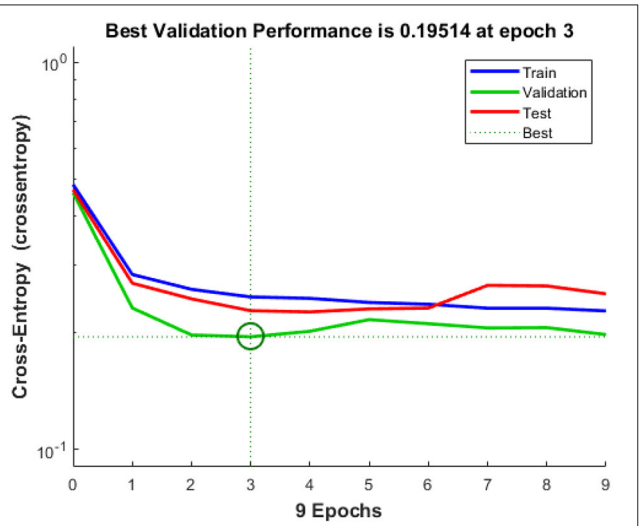


FIGURE 8 | Cross entropy.

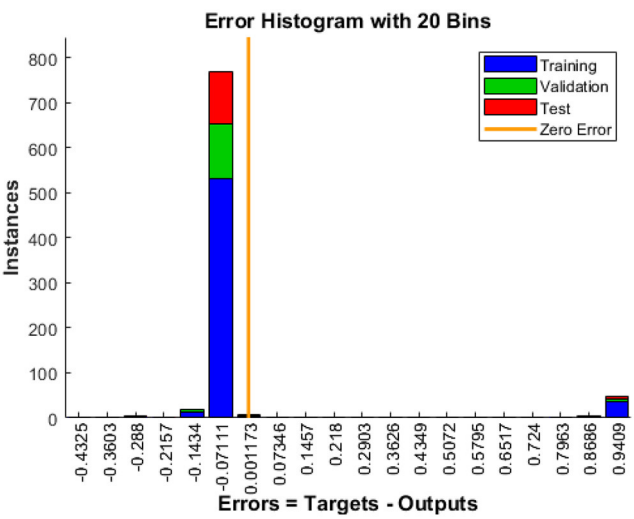


FIGURE 9 | Error histogram.

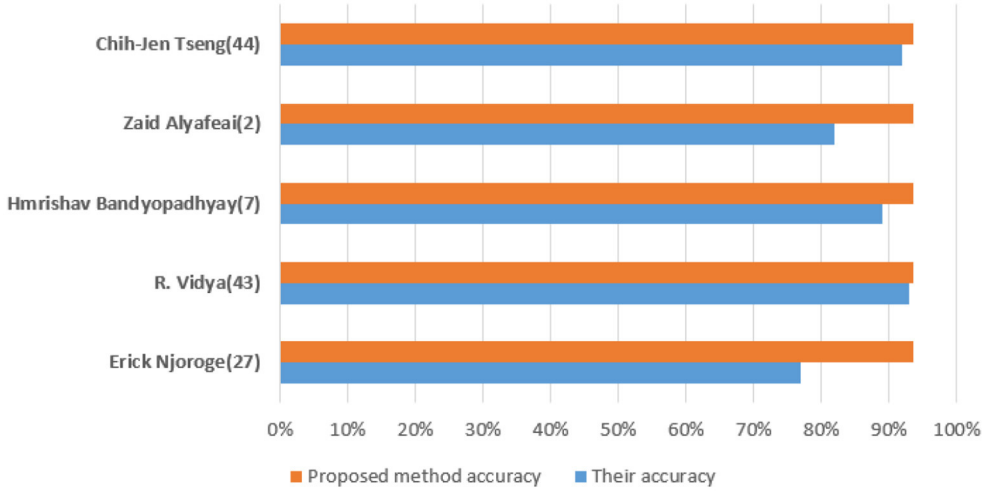


FIGURE 10 | Comparative analysis with existing works.

correctly identified over the overall patient records, as shown in Equation (19).

$$TPR = \frac{TP}{TP + FN} \tag{19}$$

False positive ratio (FPR): the ratio of the numbers of incorrectly declined records divided by the cumulative amount of total record as shown in Equation (20)

$$FPR = \frac{FP}{FP + TN} \tag{20}$$

In this study, we have proposed a new method called RF-shallow neural network to diagnose cervical cancer as the effect of the wrong diagnosis in cervical cancer or vice versa is high. Data mining provides tools and techniques to derive important data from a huge dataset by analysis. Throughout this study, ML models (RF and shallow neural network) were used for cervical cancer diagnostics to demonstrate the importance of building the model with data cleaning, the replacement of null values, and the implementation of a feature selection procedure to achieve a better accuracy prediction for an optimal function subset. This methodology used ML methodologies for cervical cancer data to assess the efficacy of classifier models by considering all medical records. Dataset missing records are filled by replacing missing values in rows with their mean and median by finding the Pearson correlation between variables. The RF's implementation as a feature selection technique and the shallow neural network training by an optimum subset of features was chosen based on the variables' value. In addition to the one target attribute biopsy for cervical cancer diagnosis, the following attributes in **Table 3** have been defined as more important features among the cervical cancer risk factors dataset.

A shallow neural network consists of one input layer, one layer of hidden neurons, and a single output layer, as shown in **Figure 5**. Neural pattern recognition uses the scaled conjugate gradient algorithm. The value of one class, "0," is treated as a negative test for a cervical cancer patient, and the value of another class, "1," is considered a positive measure for a cervical cancer patient and recommended for biopsy.

A total of 12 optimum features input space has been created after their importance and ranking. We took 70% of the data from 859 records for training, 15% for validation, and 15% for testing. To obtain the results, we used 50 hidden neurons and 1 epoch. **Table 4** shows the findings we obtained from ANN using scaled conjugate gradient using minimized cross-entropy, correct true positive rate, false-positive rate, and accuracy.

Figure 6 displays the confusion matrixes for validation, testing, and training, and the three pieces of information merged. The network findings are promising because we can conclude with many correct responses in green squares and a minimal number of wrong responses in red squares. The lower squares indicate the actual precision.

Figure 7 illustrates the receiver operating feature (ROC) curve. The colored lines in each axis represent the curves of a ROC. The ROC curve is a graphical representation of a true positive (sensitivity) vs. the false positive (1-specificity) in which the threshold has been changed. The optimal search has shown points in the upper left corner with 93% accuracy.

There have been several lines, train, validation, testing, and best in this study, as shown in **Figure 8**. In general, the ideal (dotted) line is that the other will lie on or above this (dotted) line, and we can conclude that the training has been carried out effectively. If all of the 3 (training, validation, and testing) lines cross or travel past the ideal (dotted) path, this implies that convergence has been completed. If this is not the case, retrain the network.

When the model is tested for more than 129 tests, the histogram displayed in **Figure 9** shows the error that occurs during the test process. The distribution of the error size of the histogram shows how well the neural network matches the

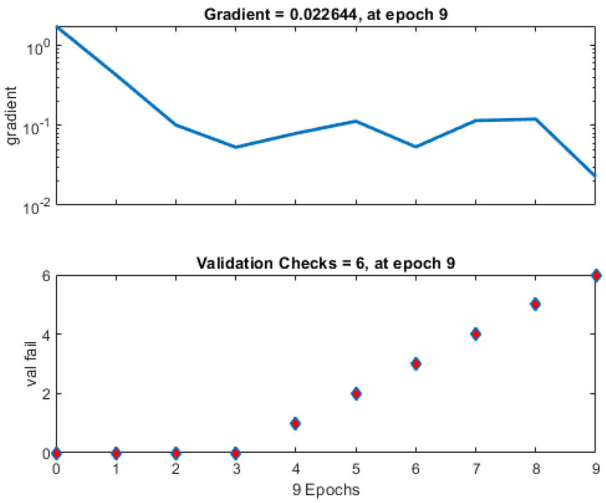


FIGURE 11 | Training state of neural network for cervical cancer diagnosis.

results. It indicates errors that the solution to mean squared error (MSE) in our scenario is good. More than 500 events are utilized for training purposes. Approximately, 129 are used for validation, and the remaining 129 is used to check which indicates an error of 0.07111.

Figure 10 shows a training state that represents the actual progress/state of the training at a given period when the training is in progress. In our situation, 6 validation errors are stated, i.e., unless 6 validation tests are performed, then training will end. Validation test error implies that the dataset has specific issues; a training algorithm cannot recognize certain instances. This indicates that due to dataset issues, validation test errors may be produced. In our case, 6 validity tests are included. As shown in **Figure 11**, modified *CervDetect* achieved 93.6% that outperforms the existing work.

6. CONCLUSION AND FUTURE WORK

This study provides a risk factor for cervical cancer. The cure percentage, the number of female cases recovered from cancer may be improved by recognizing risk factors for cervical cancer. Data mining techniques in the medical industry have an immense capacity to develop diagnostic and prognosis indicator applications that can aid in the proper initial treatment of life-threatening diseases. With various data mining techniques, it is important to enhance the standard of care in hospitals and maximize patients' recovery rates. The outcomes of the data reduction also continued to increase the efficiency of the methods. Generally, the less critical features used in the classification process, the stronger the external neural network's efficiency. In this article, we present *CervDetect*, a hybrid approach that combines RF and shallow neural network to understand the risk elements of malignant cervical formation by deploying machine learning algorithms on medical data. *CervDetect* utilizes Pearson correlation between input variables and the output variable to pre-process the data. The RF feature selection technique is used to select significant features. *CervDetect* effectively enhances the detection rate in comparison with state-of-the-art studies. Results show that the proposed approach achieved an accuracy of 93.6%, MSE error of 0.07111, FPR of 6.4%, and FNR of 100%. For future work, applying a sequential model to this dataset, such as a recursive deep neural network, can be useful for diagnosis. Due to the severity of the disease, cervical cancer is chosen as a starting point. The same work delineated here could be extended to other gynecological cancer predictions and other disease entities. Thus, the provided approach *CervDetect* showed optimal performance accuracy with the optimum features dataset.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SA and MM: conceptualization. MM: data curation. SA: formal analysis, investigation, methodology, and software. MR: funding acquisition. MG and MR: project administration, resources, and writing—review and editing. MG and SA: supervision. MM and MG: validation and visualization. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This research has been supported by the Department of Information Systems, Faculty of Management, Comenius University in Bratislava, Slovakia.

REFERENCES

1. Khamparia A, Gupta D, Rodrigues JJ, de Albuquerque VHC. DCAVN: cervical cancer prediction and classification using deep convolutional and variational autoencoder network. *Multimedia Tools Appl.* (2021) 80:30399–415. doi: 10.1007/s11042-020-09607-w
2. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput. Sci.*. (2021) 7:e390. doi: 10.7717/peerj-cs.390
3. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in Pakistan. *Comput Mater Continua.* (2021) 69:1253–69. doi: 10.32604/cmc.2021.015655
4. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
5. Rehman SU, Javed AR, Khan MU, Nazar Awan M, Farukh A, Hussien A. Personalisedcomfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents. *Enterprise Inform Syst.* (2020) 1–23. doi: 10.1080/17517575.2020.1852316

6. Javed AR, Faheem R, Asim M, Baker T, Beg MO. A smartphone sensors-based personalized human activity recognition system for sustainable smart cities. *Sustain Cities Soc.* (2021) 71:102970. doi: 10.1016/j.scs.2021.102970
7. Aslam B, Javed AR, Chakraborty C, Nebhen J, Raqib S, Rizwan M. Blockchain and ANFIS empowered IoMT application for privacy preserved contact tracing in COVID-19 pandemic. *Pers Ubiquit Comput*. (2021) 1–17. doi: 10.1007/s00779-021-01596-3
8. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst*. (2021) 1–15. doi: 10.1007/s00530-021-00774-w
9. Tripathy B, Parimala M, Reddy GT. Innovative classification, regression model for predicting various diseases. In: *Data Analytics in Biomedical Engineering and Healthcare*. Elsevier (2021). p. 179–203.
10. Javed AR, Sarwar MU, Khan S, Iwendi C, Mittal M, Kumar N. Analyzing the effectiveness and contribution of each axis of tri-axial accelerometer sensor for accurate activity recognition. *Sensors*. (2020) 20:2216. doi: 10.3390/s20082216
11. Chen H, Liu J, Wen QM, Zuo ZQ, Liu JS, Feng J, et al. CytoBrain: cervical cancer screening system based on deep learning technology. *J Comput Sci Technol*. (2021) 36:347–60. doi: 10.1007/s11390-021-0849-3
12. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfik H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. H*Hum Centric Comput Inform Sci*. (2020) 10:1–21. doi: 10.1186/s13673-020-00245-7
13. Javed AR, Fahad LG, Farhan AA, Abbas S, Srivastava G, Parizi RM, et al. Automated cognitive health assessment in smart homes using machine learning. *Sustain Cities Soc*. (2021) 65:102572. doi: 10.1016/j.scs.2020.102572
14. Sarwar MU, Javed AR. Collaborative health care plan through crowdsource data using ambient application. In: *2019 22nd International Multitopic Conference (INMIC)*. IEEE (2019). p. 1–6.
15. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*. (2005) 55:74–108. doi: 10.3322/canjclin.55.2.74
16. Salmi N, Rustam Z. Naive Bayes classifier models for predicting the colon cancer. In: *IOP Conference Series: Materials Science and Engineering. vol. 546*. IOP Publishing (2019). p. 052068.
17. Chirenje ZM, Rusakaniko S, Kirumbi L, Ngwalle EW, Makuta-Tlebere P, Kaggwa S, et al. Situation analysis for cervical cancer diagnosis and treatment in east, central and southern African countries. *Bull World Health Org*. (2001) 79:127–32.
18. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. (2015) 65:87–108. doi: 10.3322/caac.21262
19. Sarwar A, Ali M, Suri J, Sharma V. Performance evaluation of machine learning techniques for screening of cervical cancer. In: *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE (2015). p. 880–6.
20. Hajimirzaei B, Navimipour NJ. Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm. *ICT Express*. (2019) 5:56–9. doi: 10.1016/j.icte.2018.01.014
21. Prokopowicz P, Mikołajewski D, Tyburek K, Mikołajewska E. Computational gait analysis for post-stroke rehabilitation purposes using fuzzy numbers, fractal dimension and neural networks. *Bull Pol Acad Sci Techn Sci*. (2020) 191–8.
22. Bojja GR, Ofori M, Liu J, Ambati LS. Early public outlook on the coronavirus disease (COVID-19) A social media study. (2020).
23. de Brevern AG, Meyniel JP, Fairhead C, Neuvéglise C, Malpertuy A. Trends in IT innovation to build a next generation bioinformatics solution to manage and analyse biological big data produced by NGS technologies. *BioMed Res Int*. (2015) 2015. doi: 10.1155/2015/904541
24. Rehman ZU, Zia MS, Bojja GR, Yaqub M, Jinchao F, Arshid K. Texture based localization of a brain tumor from MR-images by using a machine learning approach. *Med Hypotheses*. (2020) 141:109705. doi: 10.1016/j.mehy.2020.109705
25. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
26. Lee CKH, Tse YK, Ho G, Chung SH. Uncovering insights from healthcare archives to improve operations: an association analysis for cervical cancer screening. *Technol Forecast Soc Change*. (2021) 162:120375. doi: 10.1016/j.techfore.2020.120375
27. Punjani DN, Atkotiya KH. Cervical cancer prediction using data mining. *Int J for Res Appl Sci Eng Tech*. (2017) 5:1856.
28. Breiman L. *Random Forests. Statistics Department*. Berkeley, CA: University of California (2001).
29. Sun G, Li S, Cao Y, Lang F. Cervical Cancer Diagnosis based on Random Forest. *Int J Performabil Eng*. (2017) 13:446–57. doi: 10.23940/ijpe.17.04.p12.446457
30. Mittal K, Aggarwal G, Mahajan P. Performance study of K-nearest neighbor classifier and K-means clustering for predicting the diagnostic accuracy. *Int J Inform Technol*. (2019) 11:535–40. doi: 10.1007/s41870-018-0233-x
31. Mitra P, Mitra S, Pal SK. Staging of cervical cancer with soft computing. *IEEE Trans Biomed Eng*. (2000) 47:934–40. doi: 10.1109/10.846688
32. Wu W, Zhou H. Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access*. (2017) 5:25189–95. doi: 10.1109/ACCESS.2017.2763984
33. Ashok B, Aruna P. Comparison of Feature selection methods for diagnosis of cervical cancer using SVM classifier. *Int J Eng Res Appl*. (2016) 6:94–9.
34. Tseng CJ, Lu CJ, Chang CC, Chen GD. Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Comput Appl*. (2014) 24:1311–6. doi: 10.1007/s00521-013-1359-1
35. Reif DM, Motsinger AA, McKinney BA, Crowe JE, Moore JH. Feature selection using a random forests classifier for the integrated analysis of multiple data types. In: *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*. IEEE (2006). p. 1–8.
36. Vidya R, Nasira G. Prediction of cervical cancer using hybrid induction technique: A solution for human hereditary disease patterns. *Indian J Sci Technol*. (2016) 9:1–10. doi: 10.17485/ijst/2016/v9i30/82085
37. Geeitha S, Thangamani M. Integrating HSICBFO and FWSMOTE algorithm-prediction through risk factors in cervical cancer. *J Ambient Intell Hum Comput*. (2020) 1–13. doi: 10.1007/s12652-020-02194-6
38. Anuraga G, Fernanda J, Pebrianty JW. Random forest prognostic factor in colorectal cancer. *J Phys Confer Ser*. (2019) 1217:012098. doi: 10.1088/1742-6596/1217/1/012098
39. Bandyopadhyay H, Nasipuri M. Segmentation Of Pap Smear Images For Cervical Cancer Detection. In: *2020 IEEE Calcutta Conference (CALCON)*. IEEE (2020). p. 30–3.
40. Alyafeai Z, Ghouti L. A fully-automated deep learning pipeline for cervical cancer classification. *Expert Syst Appl*. (2020) 141:112951. doi: 10.1016/j.eswa.2019.112951
41. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm. *Inform Med Unlocked*. (2019) 14:23–33. doi: 10.1016/j.imu.2019.02.001
42. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Intern Things J*. (2021). 2021:5580939. doi: 10.1155/2021/5580939
43. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inform*. (2021). doi: 10.1109/JBHI.2021.3112693
44. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an IPFS-based optimization algorithm. *J Circuits Syst Comput*. (2021) 2250010. doi: 10.1142/S0218126622500104
45. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-Technological factors affecting user's adoption of eHealth functionalities: A case study of China and Ukraine eHealth systems. *IEEE Access*. (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
46. Mohiyuddin A, Javed AR, Chakraborty C, Rizwan M, Shabbir M, Nebhen J. Secure cloud storage for medical IoT data using adaptive neuro-fuzzy inference system. *Int J Fuzzy Syst*. (2021) 1–13. doi: 10.1007/s40815-021-01104-y

47. Javed AR, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. Pp-spa: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Process Lett*. (2021) 1–18. doi: 10.1007/s11063-020-10414-5
48. Shabbir M, Shabbir A, Iwendi C, Javed AR, Rizwan M, Herencsar N, et al. Enhancing security of health information using modular encryption standard in mobile cloud computing. *IEEE Access*. (2021) 9:8820–34. doi: 10.1109/ACCESS.2021.3049564
49. Susanty M, Sahrul, Rahman AF, Normansyah MD, Irawan A. Offensive Language Detection using Artificial Neural Network. In: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. IEEE (2019) p. 350–3.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Mehmood, Rizwan, Gregus ml and Abbas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fpubh.2021.788074

Detecting COVID-19-Related Fake News Using Feature Extraction

Suleman Khan[1]*, *Saqib Hakak*[2]*, *N. Deepa*[3], *B. Prabadevi*[3], *Kapal Dev*[4] *and Silvia Trelova*[5]

[1] Air University, Islamabad, Pakistan, [2] Canadian Institute for Cybersecurity, University of New Brunswick Fredericton, Fredericton, NB, Canada, [3] School of Information Technology and Engineering, VIT University, Vellore, India, [4] Division for Institutional Planning, Evaluation and Monitoring (DIPEM), University of Johannesburg, Johannesburg, South Africa, [5] Department of Information Systems, Faculty of Management, Comenius University Bratislava, Bratislava, Slovakia

Since its emergence in December 2019, there have been numerous posts and news regarding the COVID-19 pandemic in social media, traditional print, and electronic media. These sources have information from both trusted and non-trusted medical sources. Furthermore, the news from these media are spread rapidly. Spreading a piece of deceptive information may lead to anxiety, unwanted exposure to medical remedies, tricks for digital marketing, and may lead to deadly factors. Therefore, a model for detecting fake news from the news pool is essential. In this work, the dataset which is a fusion of news related to COVID-19 that has been sourced from data from several social media and news sources is used for classification. In the first step, preprocessing is performed on the dataset to remove unwanted text, then tokenization is carried out to extract the tokens from the raw text data collected from various sources. Later, feature selection is performed to avoid the computational overhead incurred in processing all the features in the dataset. The linguistic and sentiment features are extracted for further processing. Finally, several state-of-the-art machine learning algorithms are trained to classify the COVID-19-related dataset. These algorithms are then evaluated using various metrics. The results show that the random forest classifier outperforms the other classifiers with an accuracy of 88.50%.

Keywords: COVID-19, fake news, social media, feature extraction, machine learning

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Nawab Muhammad Faseeh Qureshi,
Sungkyunkwan University,
South Korea
Wajid Rafique,
Université de Montréal, Canada

***Correspondence:**
Suleman Khan
sulemankhan@IEEE.org
Saqib Hakak
saqib.hakak@unb.ca

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *01 October 2021*
Accepted: *11 November 2021*
Published: *04 January 2022*

Citation:
Khan S, Hakak S, Deepa N, Prabadevi B, Dev K and Trelova S (2022) Detecting COVID-19-Related Fake News Using Feature Extraction. Front. Public Health 9:788074. doi: 10.3389/fpubh.2021.788074

1. INTRODUCTION

Reports about the novel coronavirus' (COVID-19) origin in Wuhan city in Hubei province, China came into the limelight in December 2019. Since then, the virus has spread in several provinces in China, and gradually to the majority of countries across the globe. Millions of people have been affected globally by this virus leading to the deaths of hundreds of thousands of people across the globe especially in countries like Italy, Spain, the United States, India, Brazil, and Russia as of June 2020 (1–3). This led the World Health Organization to declare COVID-19 as a pandemic in March 2020 (4, 5). It is estimated that on average, each person infected by COVID-19 infects around 2.5 persons if everyone goes on with their normal lives. Thus, the initial person infected may lead to 406 further infections per month. Hence, many nations have implemented lockdowns and social distancing to reduce the spread of COVID-19 (6).

Social media is flooded with millions of posts about COVID-19. Even though some of the information in social media is genuine and informative, most of the information spread in social media about COVID-19 is potentially rumor. Several doctored videos and photos about origin,

spread of the virus, vaccines, and deaths caused by COVID-19 are being shared over social media. It is estimated that around 30–35% of the news, videos, and photos spread on social media platforms are fake. This fake news travels faster than the virus itself creating widespread panic (7).

According to the International Fact Checking Network's (IFCN) study between January and April 2020, the fake news spread on social media can be categorized as follows: content about symptoms, causes, and cures, government documents, spread of the virus, misrepresentation of videos and photos, comments of politicians, and conspiracies that blame particular groups, countries, or communities for the spread of the virus. The fake news spread on social media has led to economic crisis in some countries. For instance, in some countries people stopped consuming non-vegetarian food as fake news was spread that animals and birds could be infected with COVID-19 and consuming non-vegetarian food may spread the virus in people. This had a severe impact on the sales of non-vegetarian food in some countries affecting the livelihood of many people (8). **Figure 1** depicts the application of machine learning (ML) algorithms (9, 10) to detect fake news spread on social media.

In view of the serious impacts of the spreading of fake news which affects the privacy and security (11, 12) of users, the need of the hour is to design a successful mechanism to detect/predict fake news. In this work, an attempt is made to classify the fake news regarding COVID-19. The dataset used in this work is a fusion of fake and real news collected from several social media platforms and websites (13). This dataset is then subjected to preprocessing to remove some irrelevant text such as URL punctuations and noisy data, etc. Then, the resultant text is split into several small words using tokenization. Then, the most important features are extracted from these tokens. These features are then trained by state-of-the-art ML algorithms to classify whether the COVID-19-related news is fake or real.

The main contributions of the current work are as follows:

- Classification of a novel COVID-19 fake news dataset collected from several social media platforms.
- Generating 39 linguistic and sentiment (14–16) features from text.
- Extracting the most important features from the COVID dataset to improve the accuracy of the ML algorithms.
- An accuracy of 88.50% is achieved by the random forest classifier.

The rest of the paper is organized as follows: In section 2, related works on COVID-19 fake news are discussed. ML algorithms used in this work to classify the fake news dataset are discussed in section 3. The proposed architecture is presented in section 4. The experiment results are presented along with the discussion in section 5. Finally the paper is concluded in section 6.

2. LITERATURE SURVEY

Artificial intelligence (AI) has made a significant contribution to different aspects of the COVID-19 outbreak. A survey on various applications of AI for fighting COVID was presented in Nguyen (17). Specifically, deep learning algorithms play a vital role in empowering AI in most of the applications with larger datasets with different types of data. Some of the applications are computer-generated for predicting infected people from a large gathering through a high-resolution image. The most prominent application is in text mining assisted with natural language processing (NLP) for extracting semantically related information from numerous texts in multilingual news data collected from various social networking platforms. The details obtained from these models such as susceptible infected recovered (SIR) models can be injected into deep learning models for the prediction of COVID-19 trends of transmission and can avoid future pandemic or epidemic diseases (18). Furthermore, a mapping study on various AI models with big data for detecting fake news on social media was carried out in Meneses Silva et al. (19). They suggested that the long short term memory model provides better results than other algorithms but incurs more processing power.

Although social media is the platform that spreads news faster with low cost than any other media among all age groups, it is progressively more dangerous in spreading fake news. Many research projects were conducted on detecting fake news from social media. A two-step model for detecting fake news from the real story in social media using artificially intelligent algorithms was designed (20). Three real datasets were considered for the study. The dataset with unstructured data was preprocessed to obtain meaningful data and was represented using vectors. In turn, the results were generated for 23 supervised learning algorithms and compared based on four evaluation metrics. On the other hand, a biostatistical analysis of the COVID-19 pandemic was carried out in Bandyopadhyay and Dutta (21) using the KNN classifier. They collected data based on certain information from the news topics in social media through multi-document summarization. This summarization method extracted the information based on the lexical information of the topic and social patterns. The KNN classifier predicted fake news with an accuracy of 80%.

Furthermore, the study in Groza (22) investigated the spread of deceptive information on social media through COVID Ontology. The reasoning in the natural language is converted into description logic for perceiving inconsistencies among different medical sources. The authors argue for an ontology-based approach for detecting fake news related to the coronavirus. The proposed model explores how well description logics (DLs) understand contradictions in figuring out whether they are true or false. The authors pay particular attention to translating the natural language into DL using the FRED translator, after which, justification is performed to test the performance of the prediction using the racer method. The proposed ontology model reveals misconceptions that circulate in digital news and shows accurate facts. However, potential research can be expanded by program evaluation and provides a verbal remedy for any conflict that exists.

Fake news or false information about this COVID-19 pandemic are very prominent and should be detected to avoid unwanted chaos. A FakeCovid dataset from a multilingual news article was generated, and a classifier was designed to detect the fake news from the dataset generated (23). BERT, the ML-based

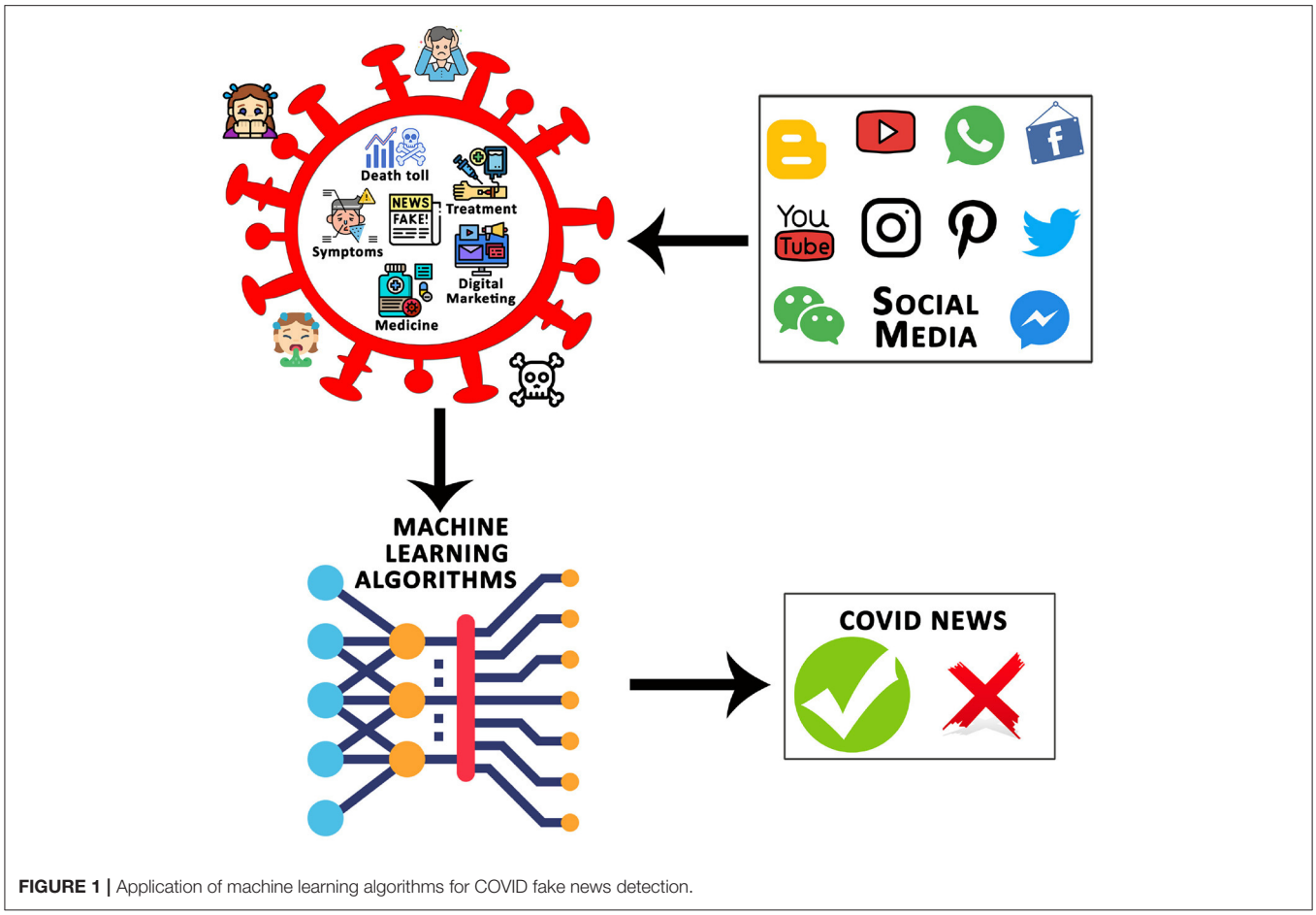


FIGURE 1 | Application of machine learning algorithms for COVID fake news detection.

classifier model, performed better on the dataset generated with an F1 score of 0.76 at the initial stage of COVID 19. The model does not have a polyglot, so they annotated categories manually for three languages.

The study in Daley (24) used ML algorithms for the automatic classification of fake news on COVID-19. The features considered for evaluation included the count of motion words and relativity words, prepositions in the headlines of the news website, tone expressed, and word count. Also, any news media would concentrate on the increased count of these parameters to convince the customers on the fake news outlets. They attained 79% accuracy on 5-fold cross-validation using a decision tree classifier, which outperformed other algorithms in the scikit-learn API. An automated framework for detecting COVID-19 cases from X-ray images of the raw chest was proposed in Ozturk et al. (25). They used a deep neural network, namely the DarkCovidNet model (used for real-time object detection) with 17 convolution layers, each with different filters. The average age of the positive COVID patients in the dataset was 55. The model was used for binary (COVID or not) and multiclass classification (COVID or pneumonia or not) with an accuracy of 98 and 87%, respectively. The performance of the system was assessed with radiologists only, and the robustness of the system can be tested with large datasets.

Therefore, it is evident from the survey that very few attempts were carried out for detecting the fake news in COVID. Although various deep learning techniques were employed for detection, the datasets taken for analysis are very small. It is mandatory to have a strong model to detect false information on COVID-19 to avoid unwanted chaos in different aspects of COVID-19.

3. PROPOSED ARCHITECTURE

In this section, the proposed work is discussed in detail.

The steps involved in the proposed work are shown in **Figure 2** and are summarized as follows:

- The dataset used in this study is collected from various websites and social media cites.
- Data preprocessing is performed on the dataset by removing URLs, punctuation marks, and empty columns.
- After text preprocessing, tokenization is performed to convert the larger text into words or small lines.
- In the next step, extracting the features from text is performed.
- The features extracted are then passed on to the state-of-the-art ML learning algorithms to train the model.
- Various evaluation metrics are used to evaluate the performance of the ML algorithms.

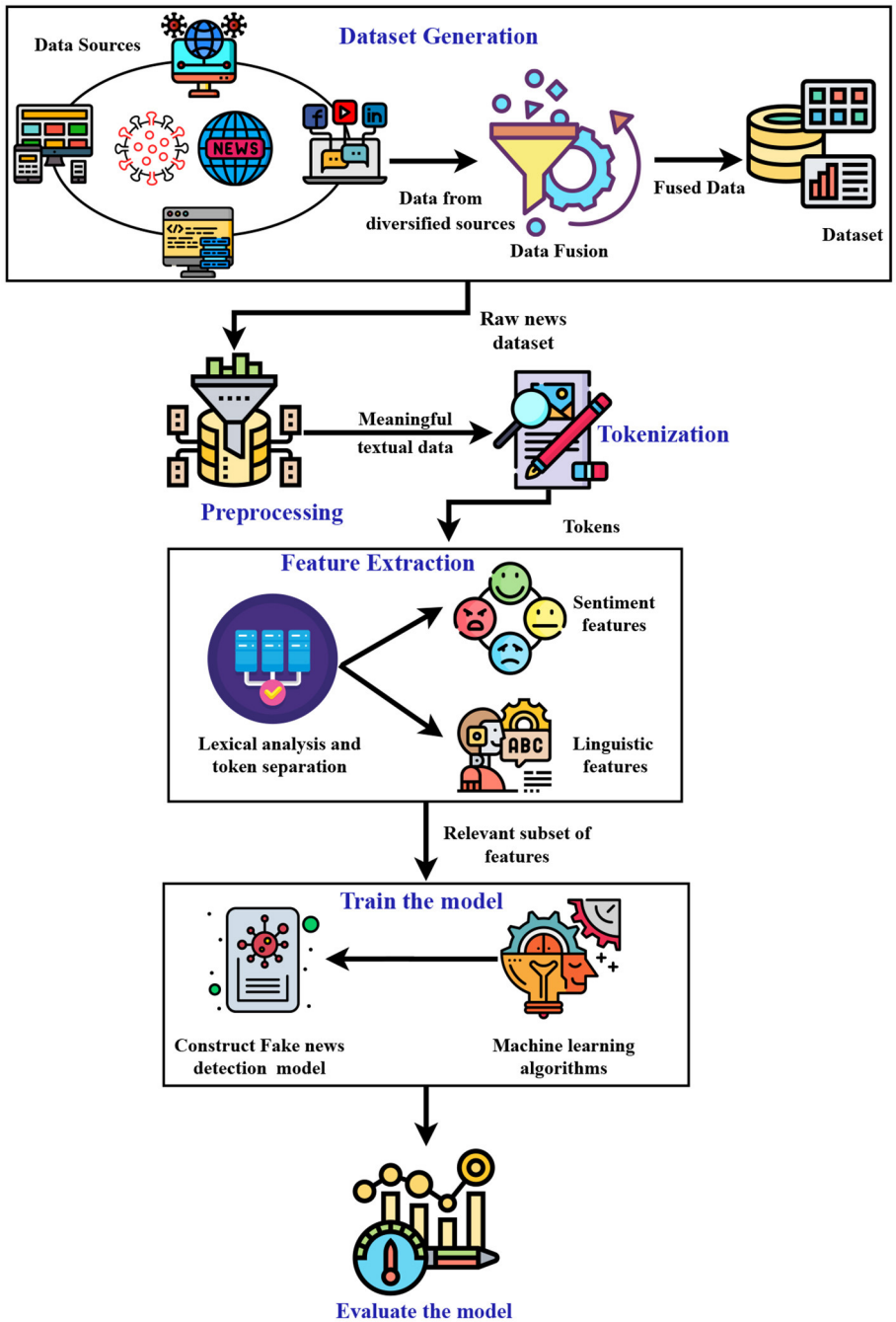


FIGURE 2 | Proposed architecture.

The rest of this section discusses various steps involved in the proposed work.

3.1. Dataset Description

The dataset used in this work is a fusion of several fake and real news articles about COVID-19 which are collected across several platforms such as Facebook, Twitter, The New York Times, Harvard Health Publishing, WHO, etc. The dataset has 1,164 instances out of which 586 instances are true and the remaining 578 are fake news (26).

3.2. Preprocessing and Tokenization

The dataset considered for this work is clean. However, some unnecessary symbols which have an impact on the final classification of the news are to be removed from the dataset. To remove the unnecessary symbols, such as

punctuation marks, URLs are removed from the dataset as part of preprocessing.

Tokenization is the process of splitting text into a set of tokens. The fake news dataset is tokenized to convert the long sentences into small words/tokens.

3.3. Feature Extraction

The major contribution of this work is the extraction of important features from the COVID-19 fake news dataset. Feature extraction plays a very important role in text processing as it reduces the dimension of feature space by considering only the important features (27–29). To extract the features, the named-entity recognition (NER) approach is used in our work. The NER is a popular approach for feature extraction that can classify unstructured text based on location, person names, quantities, etc. (30).

In this study, 39 features are created from the COVID-19-related fake news dataset. The extracted features are represented in **Table 1**.

3.4. Training of the Dataset Using ML Algorithms

In the next step, the extracted features are trained using several state-of-the-art ML algorithms. The ML algorithms used in this work are linear SVC, logistic regression, SGD classifier, random forest classifier, bagging classifier, AdaBoost classifier, decision tree classifier, and K-nearest neighbors classifier (KNN) (31). These algorithms are explained briefly in the rest of this section.

TABLE 1 | Features extracted from text.

Feature name	Data type	Feature name	Data type
News source	Non-Numeric	Num of?	Numeric
Num of Stopwords	Numeric	Num of /	Numeric
Num of @	Numeric	Num of #	Numeric
Num of numeric values	Numeric	Num of uppercase characters	Numeric
Num of lowercase characters	Numeric	Num of all uppercase characters	Numeric
Text language	Numeric	Word count	Numeric
Character count	Numeric	Sentence count	Numeric
Average word length	Numeric	Average sentence length	Numeric
Positive Sentiment Score	Numeric	Negative Sentiment Score	Numeric
Neutral Sentiment Score	Numeric	Compound Sentiment Score	Numeric
Person	Numeric	NORP	Numeric
FAC	Numeric	Organization	Numeric
GPE	Numeric	Location	Numeric
Product	Numeric	Event	Numeric
Work of Art	Numeric	Law	Numeric
Language	Numeric	Date	Numeric
Time	Numeric	Percent	Numeric
Money	Numeric	Quantity	Numeric
Cardinal	Numeric	Ordinal	Numeric
Text Polarity	Numeric		

3.4.1. Random Forest Algorithm

The random forest classifier is one of the supervised ML algorithms used for classification and regression purposes (32). The random forest method constructs decision trees from datasets and creates a forest made of trees. Then it collects the prediction value from each decision tree and computes the optimal solution with the help of voting. It is used as an ensemble method and helps to avoid overfitting (33). The following steps are implemented in the random forest classifier:

1. Random data samples are selected from the available dataset.
2. A decision tree is constructed for each data sample selected, and a prediction value is obtained from each decision tree. The Gini coefficient method is applied for node splitting as follows:

$$Gini(D) = 1 - \sum_{i=1}^{n} P_i^2 \quad (1)$$

   where D is the sample dataset and P_i is the probability of decision class that can appear in D
3. A voting method is applied on the prediction values obtained from each decision tee.
4. A final prediction result is chosen from the prediction value with more votes.

3.4.2. AdaBoost Classifier

The AdaBoost classifier (34) is an ensemble boosting classification method which combines several weak classifiers to build a strong classifier. The weak classifier performs the

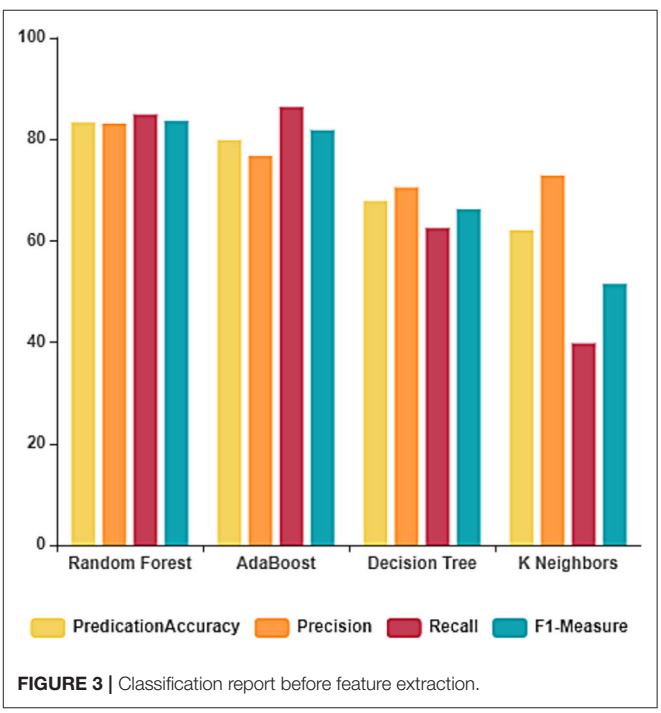


FIGURE 3 | Classification report before feature extraction.

TABLE 2 | Performance of the ML algorithms before feature extraction.

Algorithm	Predication accuracy	Precision	Recall	F1-measure	Predication time (S)	Training time (S)
Random forest classifier	83.33	83.14	84.90	83.61	0.065862	0.383069
AdaBoost classifier	79.88	76.76	86.36	81.82	0.056415	0.347593
Decision tree classifier	67.81	70.51	62.50	66.26	0.003247	0.091995
K-nearest neighbor classifier	62.06	72.91	39.77	51.47	0.084975	0.001555

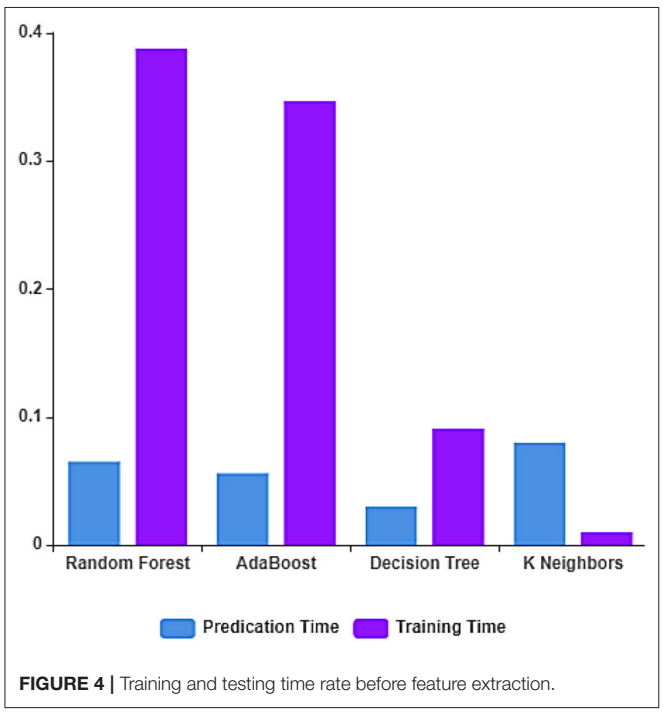


FIGURE 4 | Training and testing time rate before feature extraction.

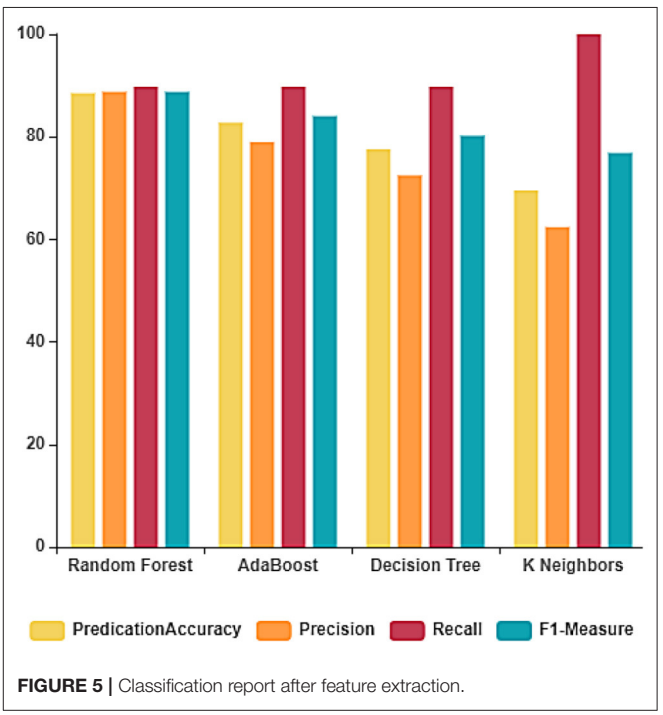


FIGURE 5 | Classification report after feature extraction.

classification of one dimension of the input vector. The number of weak classifiers are increased during the training process to obtain more accuracy. The classifier applies minimum error to compute the weight of the newly added weak classifier. Then it updates the weights of each training dataset and sends the value to the recently added weak classifier. The selection of weak classifiers is very important for this algorithm. The result of the strong classifier $S(f)$ is represented in the equation given below (35):

$$S(f) = sign\left\{\sum_{t-1}^{T} \beta_t h_t(f)\right\} \quad (2)$$

where f denotes the input feature vector, T denotes the number of input vectors, $h_t(f)$ denotes weak classifiers, and β_t denotes the weights of weak classifiers.

3.4.3. Decision Tree Classifier

The decision tree classifier is one of the more popular ML algorithms applied for classification and prediction problems on supervised data. It classifies the trained dataset into trees and rules. It defines methods for the classification of categorical data characterized by their factors. It works better on classification problems where extensive data are involved, and is therefore used in many data mining applications. The first step in the decision tree classifier is to find the attribute to split the tree. Entropy measure is used to identify the best attribute as the root of the decision tree. The entropy measure is used to calculate the update in homogeneity that results from the split on every attribute. Entropy D of the dataset is defined as follows:

$$Entropy(D) = \sum_{i=1}^{n} -p_i log_2(p_i) \quad (3)$$

Where D is the data partition, n defines the number of decision classes, and p_i defines the proportion of the records that comes under the decision class i. Information gain is used for this calculation. This is defined as the difference between the entropy of the dataset calculated before the split (D1) and the entropy calculated from the partitioned split (D2). The formula to calculate the information gain is given as follows:

$$InformationGain(G) = Entropy(D_i) - Entropy(D_2) \quad (4)$$

The attribute with maximum information gain is selected to build the tree with a root node. A decision tree is constructed recursively using the obtained data partition until there are no attributes left. And decision rules can be defined from the constructed decision tree.

3.4.4. K-Nearest Neighbor Classifier

The K-nearest neighbor (KNN) (36) classifier is one of the popularly used supervised learning algorithms for classification and regression. It is a non-parametric and lazy algorithm in which the method is defined from the dataset. Also it does not require a training dataset for the development of the classification model. K is the number of nearest neighbors and it is the deciding factor. The classifier is based on the estimation of the nearest neighbor. It works on the calculation of a similarity value defined by the distance measure. The training step stores all the instances and their corresponding class labels. The test phase for a new instance 's' is calculated using the following steps:

1. Calculate the distance between the instance 's' in the test data and every instance in the training dataset 'd' using the Euclidean distance formula given as follows:

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2)} \tag{5}$$

   where n is the total number of instances in the dataset, x_i represents instances in the training dataset, and y_i are instances in the test dataset 'd'.
2. Arrange the distances in ascending order and select the first 'k' instances.
3. Find the most frequent decision class in the calculated 'k' nearest neighbors.

4. EXPERIMENT RESULTS

The experiments on the COVID-19 fake news dataset are carried out in Google Colab, an open source cloud-based graphical processing unit (GPU)-based platform offered by Google Inc. The programming language used to implement the ML algorithms is Python 3.7.

A total of 70% of the COVID-19 fake news dataset is used for training the ML algorithms and the other 30% of the dataset is used for validating and testing the ML algorithms. The performance of the ML algorithms is evaluated on several metrics like accuracy, precision, recall, F1-measure, prediction time, and training time.

The rest of this section discusses the results obtained by the ML algorithms on the COVID-19 fake news dataset before and after the feature extraction.

4.1. COVID-19 Fake News Experiment Results Before Feature Extraction

The performance of the ML algorithms on the COVID-19 fake news dataset before the feature extraction is depicted in **Figure 3** and **Table 2**. From the table, it is evident that the random forest classifier achieves better prediction accuracy, precision, recall, and F1-measure with 83.33, 83.14, 84.90, and 83.61%, respectively. The random forest classifier trains the dataset in the least time when compared to other ML algorithms. The training and testing time of the ML algorithms on the dataset considered before feature extraction is depicted in **Figure 4**.

4.2. COVID-19 Fake News Experiment Results After Feature Extraction

The performance of the ML algorithms on the COVID-19 fake news dataset after the feature extraction is depicted in **Figure 5** and **Table 3**. From these results, it is evident that the random

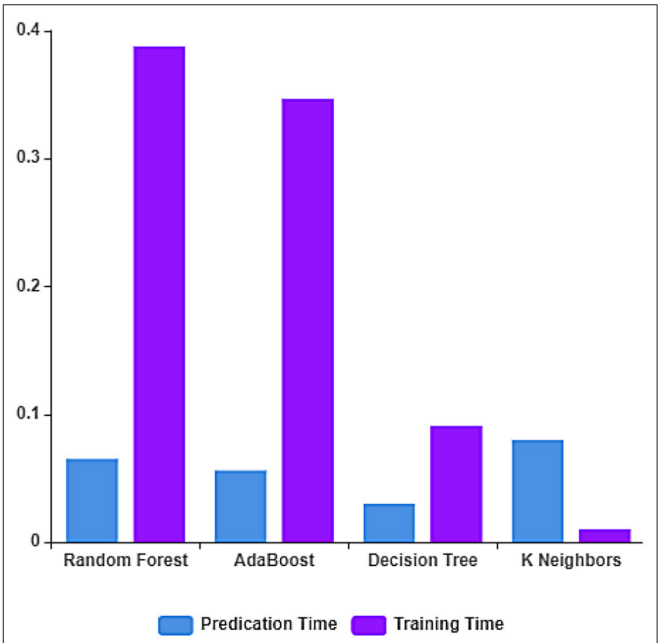


FIGURE 6 | Training and Testing time rate after feature extraction.

TABLE 3 | Performance of the ML algorithms after feature extraction.

Algorithm	Predication accuracy	Precision	Recall	F1-measure	Predication time (S)	Training time (S)
Random forest classifier	88.50	87.77	89.77	88.76	0.004685	0.011387
AdaBoost classifier	82.75	79.00	89.77	84.04	0.004750	0.053603
Decision tree classifier	77.58	72.47	89.77	80.20	0.002851	0.081209
K-nearest neighbor classifier	69.54	62.41	100	76.85	0.003293	0.008997

forest classifier outperforms the other classifiers in terms of accuracy and precision with 88.50 and 87.77%, respectively. Whereas, linear SVC achieved better recall and F1-measure with 89.77 and 88.76%, respectively.

The training and prediction time of the ML algorithms on the dataset considered after feature extraction is depicted in **Figure 6**. From the figure, it can be observed that the decision tree classifier has the least prediction time and the linear SVC classifier has the least training time.

5. DISCUSSION

From the results section, it is clear that the performance of the machine learning algorithms improves after they are trained with extracted features from the COVID-19 fake news dataset. Since the size of the dataset is approximately 1,100 records, ML algorithms are chosen for classification rather than deep neural network-based algorithms.

The raw fake news dataset has several words that have no affect on the classification results. When the ML algorithms are trained by the raw dataset without feature extraction, there is a very high chance that the performance of the ML algorithms will be affected by some of the frequent words in the text that have no effect on the classification results. Hence, in this work, we employed NER to extract features from the fake news dataset. The extracted features are then fed to the ML algorithms for training. The comparison between the results of the ML algorithms before and after the feature extraction prove that the performance of the ML algorithms increase after feature extraction.

6. CONCLUSION

In this work, a sophisticated model that extracts important features from a COVID-19 fake news dataset is presented that improves the performance of ML algorithms. Social media applications are the most significant source to spread any kind of information. So in his work, the dataset related to COVID-19 which was the fusion of several diversified social media platforms and news websites is used for evaluating the performance of the proposed model. A data preprocessing step is applied on the dataset to remove several unnecessary symbols like tags and URLs from the dataset. Later on, to enhance the performance of the machine learning algorithms, the relevant subset of features categorized into linguistic and sentiment features are extracted using the NER approach. The extracted features are trained using several state-of-the-art ML algorithms, namely the random forest classifier, AdaBoost classifier, decision tree classifier, and KNN classifier. The performance of the ML algorithms is analyzed before and after the feature extraction. Random forest provides better accuracy at 83.33% without feature extraction, whereas KNN trains the dataset faster than others. Of several ML algorithms, the random forest classifier with feature extraction outperforms the others with an accuracy of 88.50%, precision of 87.77%, recall of 89.77%, and F1 score of 88.76%. The experiment results prove the importance of feature extraction in fake news detection.

Further in future, audio clips, images, and video clips can be considered for generating fake news data as they are the most attractive way to spread fake information. Deep learning algorithms and natural language processing for the semantical understanding of linguistic features can be further analyzed on a dataset containing text, images, audio, and video.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.788074/full#supplementary-material

REFERENCES

1. Livingston E, Bucher K. Coronavirus disease 2019 (COVID-19) in Italy. *JAMA*. (2020) 323:1335–5. doi: 10.1001/jama.2020.4344
2. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
3. Srivastava G. *The Impact of the COVID-19 Pandemic on Mental Health of Children and Adolescents*. (2020). Available online at: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-414191
4. Pham QV, Nguyen DC, Hwang WJ, Pathirana PN. Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: a survey on the state-of-the-arts. *IEEE Access*. (2020) 8:130820–39. doi: 10.20944/preprints202004.0383.v1
5. Hakak S, Khan WZ, Imran M, Choo KKR, Shoaib M. Have you been a victim of COVID-19-related Cyber incidents? Survey, taxonomy, and mitigation strategies. *IEEE Access*. (2020) 8:124134–44. doi: 10.1109/ACCESS.2020.3006172
6. Organization WH. Coronavirus disease 2019 (COVID-19): situation report, 72. (2020).
7. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*. (2020) 31:770–80. doi: 10.1177/0956797620939054
8. Brennen JS, Simon F, Howard PN, Nielsen RK. Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*. (2020) 7:1. Availble Online at: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation
9. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access*. (2020) 8:28462–74. doi: 10.1109/ACCESS.2020.2968537

10. Gulati A, Aujla GS, Chaudhary R, Kumar N, Obaidat MS. Deep learning-based content centric data dissemination scheme for internet of vehicles. In: *2018 IEEE International Conference on Communications (ICC)*. Kansas City, MO: IEEE (2018). p. 1–6.
11. Azab A, Layton R, Alazab M, Oliver J. Mining malware to detect variants. In: *2014 Fifth Cybercrime and Trustworthy Computing Conference*. Auckland: IEEE (2014). p. 44–53.
12. Alazab M, Layton R, Broadhurst R, Bouhours B. Malicious spam emails developments and authorship attribution. In: *2013 Fourth Cybercrime and Trustworthy Computing Workshop*. Sydney, NSW:IEEE (2013). p. 58–68.
13. Li S. *Explore COVID-19 Infodemic*. (2020). Available online at: https://towardsdatascience.com/explore-covid-19-infodemic-2d1ceaae2306
14. Gaddekallu T, Soni A, Sarkar D, Kuruva L. Application of sentiment analysis in movie reviews in *Sentiment Analysis and Knowledge Discovery in Contemporary Business*. IGI Global (2019), 77–90. doi: 10.4018/978-1-5225-4999-4
15. Asghar MZ, Subhan F, Ahmad H, Khan WZ, Hakak S, Gadekallu TR, et al. Senti-eSystem: a sentiment-based eSystem-using hybridized fuzzy and deep neural network for measuring customer satisfaction. *Softw Pract Exp*. (2020) 51:571–59. doi: 10.1002/spe.2853
16. Maddikunta PKR, Gadekallu TR, Al-Ahmari A, Abidi MH, et al. Location based business recommendation using spatial demand. *Sustainability*. (2020) 12:4124. doi: 10.3390/su12104124
17. Nguyen TT. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. (2020). doi: 10.36227/techrxiv.12743933
18. Du S, Wang J, Zhang H, Cui W, Kang Z, Yang T, et al. Predicting COVID-19 using hybrid AI model. (2020). doi: 10.2139/ssrn.3555202
19. Meneses Silva CV, Silva Fontes R, Colaço Júnior M. Intelligent fake news detection: a systematic mapping. *J Appl Security Res*. (2020) 16:1–22. doi: 10.1080/19361610.2020.1761224
20. Ozbay FA, Alatas B. Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A*. (2020) 540:123174. doi: 10.1016/j.physa.2019.123174
21. Bandyopadhyay S, DUTTA S. The analysis of fake news in social medias for four months during lockdown in COVID-19-a study: biostatistical analysis of COVID-19. *Xeno J Biomed Sci*. (2020) 1:1–6. doi: 10.20944/preprints202006.0243.v1
22. Groza A. Detecting fake news for the new coronavirus by reasoning on the Covid-19 ontology. *arXiv preprint* arXiv:200412330. (2020).
23. Shahi GK, Nandini D. FakeCovid-a multilingual cross-domain fact check news dataset for COVID-19. *arXiv preprint* arXiv:200611343. (2020).
24. Daley BP. Leveraging machine learning for automatically classifying fake news in the covid-19 outbreak. (2020).
25. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med*. (2020) 121: 103792. doi: 10.1016/j.compbiomed.2020.103792
26. Github. *Fake News Dataset*. (2020). Available online at: https://raw.githubusercontent.com/susanli2016/NLP-with-Python/master/data/corona_fake.csv (accessed on July 14, 2020).
27. Reddy GT, Reddy MPK, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G, et al. Analysis of dimensionality reduction techniques on big data. *IEEE Access*. (2020) 8:54776–88. doi: 10.1109/ACCESS.2020.2980942
28. Gadekallu TR, Rajput DS, Reddy MPK, Lakshmanna K, Bhattacharya S, Singh S, et al. A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *J Real Time Image Proc*. (2020) 18:1383–96. doi: 10.1007/s11554-020-00987-8
29. Rajput DS, Singh AP, Agarwal Y, Reddy PK, Reddy GT. Feature selection analysis for multimedia event detection in *IOP Conference Series: Materials Science and Engineering,* IOP Publishing (2017) 263:042004.
30. Li J, Sun A, Han J, Li C. A survey on deep learning for named entity recognition in *IEEE Transactions on Knowledge and Data Engineering* (IEEE) (2020) *arXiv [Preprint]*. arXiv: 1812.09449. doi: 10.1109/TKDE.2020.2981314
31. Patel H, Singh Rajput D, Thippa Reddy G, Iwendi C, Kashif Bashir A, Jo O. A review on classification of imbalanced data for wireless sensor networks. *Int J Distribut Sensor Netw*. (2020) 16:1550147720916404. doi: 10.1177/1550147720916404
32. Maddikunta PKR, Srivastava G, Gadekallu TR, Deepa N, Boopathy P. Predictive model for battery life in IoT networks. *IET Intell Transp Syst*. (2020) 14:1388–95.
33. Yan X, Jin Y, Xu Y, Li R. Wind turbine generator fault detection based on multi-layer neural network and random forest algorithm. In: *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. Chengdu: IEEE (2019). p. 4132–6.
34. Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. In: *European Conference on Computational Learning Theory*. Springer (1995). p. 23–37.
35. Cheng WC, Jhan DM. Triaxial accelerometer-based fall detection method using a self-constructing cascade-AdaBoost-SVM classifier. *IEEE J Biomed Health Inform*. (2012) 17:411–9. doi: 10.1109/JBHI.2012.2237034
36. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inform Theory*. (1967) 13:21–7. doi: 10.1109/TIT.1967.1053964

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Khan, Hakak, Deepa, Prabadevi, Dev and Trelova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fpubh.2021.813717

Analysis of Tracheobronchial Diverticula Based on Semantic Segmentation of CT Images *via* the Dual-Channel Attention Network

Maoyi Zhang[1], *Changqing Ding*[2]* *and Shuli Guo*[1]

[1] School of Control Science and Engineering, Beijing Institute of Technology, Beijing, China, [2] Department of Imaging, Fengxian People's Hospital, Xuzhou, China

Tracheobronchial diverticula (TD) is a common cystic lesion that can be easily neglected; hence accurate and rapid identification is critical for later diagnosis. There is a strong need to automate this diagnostic process because traditional manual observations are time-consuming and laborious. However, most studies have only focused on the case report or listed the relationship between the disease and other physiological indicators, but a few have adopted advanced technologies such as deep learning for automated identification and diagnosis. To fill this gap, this study interpreted TD recognition as semantic segmentation and proposed a novel attention-based network for TD semantic segmentation. Since the area of TD lesion is small and similar to surrounding organs, we designed the atrous spatial pyramid pooling (ASPP) and attention mechanisms, which can efficiently complete the segmentation of TD with robust results. The proposed attention model can selectively gather features from different branches according to the amount of information they contain. Besides, to the best of our knowledge, no public research data is available yet. For efficient network training, we constructed a data set containing 218 TD and related ground truth (GT). We evaluated different models based on the proposed data set, among which the highest MIOU can reach 0.92. The experiments show that our model can outperform state-of-the-art methods, indicating that the deep learning method has great potential for TD recognition.

Keywords: tracheobronchial diverticula identification, semantic segmentation, deep learning, dual-channel, attention

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Sharnil Pandya, Symbiosis International University, India
Rutvij Jhaveri, Pandit Deendayal Petroleum University, India
Mohammad Kamrul Hasan, National University of Malaysia, Malaysia

***Correspondence:**
Changqing Ding
dcq2020@126.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *12 November 2021*
Accepted: *06 December 2021*
Published: *05 January 2022*

Citation:
Zhang M, Ding C and Guo S (2022) Analysis of Tracheobronchial Diverticula Based on Semantic Segmentation of CT Images via the Dual-Channel Attention Network. Front. Public Health 9:813717. doi: 10.3389/fpubh.2021.813717

INTRODUCTION

Background

Tracheobronchial diverticula (TD), or paratracheal air cystic (PTAC), is defined as a congenital or acquired air cyst located in a lateral wall of trachea (1). Specifically, it is a cystic lesion involving primarily the outpouching of the trachea and main bronchial lumen and is located mainly on small round air cysts lesion beside the superior mediastinal trachea.

Tracheobronchial diverticula, mostly discovered in the chest CT by accident, is usually asymptomatic. Chronic cough, dyspnea, stridor or recurrent tracheitis, dysphagia, dysphasia, odynophagia, neck pain, hoarseness, suffocation, and hemoptysis caused by TD are less frequent. In rare cases, TD may lead to the chronic infection of the tracheobronchus. Only when the relationship

between TD and symptoms is confirmed can the symptomatic treatment of antibiotics, mucus solubilizers, and physical therapies be used. The endoscope check includes electrolumes and laser or electroconding endoscopes. Inspections including detachable and longitudinal frame suction and biopsy, thoracoscope, braking, and longitudinal isolation are only needed in rare cases. Before the advent of the CT, due to insufficient understanding of this disease, the lack of specificity of the clinical symptoms, and the restrictions of imaging methods such as chest portions, only severe symptoms could be diagnosed. Thanks to the wide application of multislice spiral CT and the advancement of computer technology, the detection rates of the tracheal and bronchial diverticulums are greatly improved. The coronary and sagittal images are useful in showing the relationship between the tracheal cavity and TD.

Tracheobronchial diverticula is mainly identified by the following lesions: apical lung hernia, trachea laceration, paramediastinal emphysema, esophageal diverticulum, paratracheal air cyst, pneumomediastinum, subcutaneous cervical emphysema, mediastinal emphysema, cervical subcutaneous emphysema, and cervical tracheal duplication. CT should be taken as the main diagnostic tool. The whole picture of the diverticulum and its communication with trachea and bronchi can be observed through thin-section CT and its 3D reconstruction; so we should give full play to spiral CT's advantages in diagnosing tracheal diverticulum, and avoid unnecessary invasive inspection. Bronchoscopy is another tool for diagnosis, but with it, some patients suffering from tracheal occlusion can be misdiagnosed as narrow neck TD or fibrous connection. Unfortunately, not all physicians have sufficient understandings in this regard, as a large number of missed diagnoses still occur in many hospitals worldwide. In addition, since the lesion area of TD is small it poses great challenges to diagnosis. Therefore, designing an efficient and accurate automated method to detect TD can significantly reduce the workload and the missed diagnosis or misdiagnosis rates, ultimately enhancing the diagnostic effectiveness. Once the method is widely applied to hospitals, large sets of diagnostic data obtained can be utilized to explore the potential causes further, and to determine the prevalence rate, morphology (i.e., quantity, size), and clinical features of TD. For a clearer illustration, **Figure 1** shows the scenarios of TD's application.

Related Work

The related work is divided into four parts: tracheobronchial diverticula, machine learning, deep learning, and cloud technology.

Research Status of TD

Few cases of TD have been reported since the first description by Rokitansky in 1838 (2). With the increasingly wide application of CT as a high-performance and non-destructive testing method in the medical field, especially as the main screening method for early lung cancer, the detection rate of PTAC has increased significantly by about 4–8.1% (3–5), which is 1% higher than previously reported postmortem prevalence (0.3% in children) (6). Pace et al. suggested that this disease was associated with the congenital defect or abnormal development of the posterior tracheal membrane, or possibly with acquired chronic obstructive pulmonary disease (7). Ahmet et al. used chest multidetector computed tomography (MDCT) to determine the prevalence of paratracheal air cysts and their correlation with different lung diseases, and finally proved that air cysts are associated with chronic obstructive pulmonary disease (3).

Unlike traditional medical segmentation, the lesion area of TD is relatively small in both real conditions and CT image. In the medical field, most researchers only reported related cases and gave prescriptions but did not design an automatic method for detection.

Research Status of Machine Learning

In the field of machine learning, Deepa et al. (8) used Ridge Adaline Stochastic Gradient Descent Classifier for the early prognosis of diabetes, with higher accuracy than other algorithms. Dhanamjayulu et al. (9) adopted machine learning algorithms to identify malnutrition from facial images. Ghazal et al. reviewed the machine learning in smart healthcare (10) and an improved SVM method for diagnosing Hepatitis C (11). Shah et al. (12) proposed two gender-based risk classification and age-based risk classification methods to identify cardiovascular diseases (CVD). The accuracy of the model can reach 98%, which improved the identification efficiency. In addition, the progress of machine learning in the direction of medical data mining (13) is also gratifying.

Research Status of Deep Learning

In recent years, deep learning (14) has shown a booming trend of development, in which many networks for medical segmentation of small and medium targets have emerged. For example, U-net (15) is specially developed for medical image segmentation. Its U-shaped structure is an asymmetrical structure of up- and downsampling, whose feature mapping can be connected through the intermediate jump connection, thus enriching the extracted semantic information. Shrivastava et al. (16) presented a learning mechanism method for cognitive maps, which has been successfully applied to identify problems in targets. Jégou et al. (17) restored the complete input resolution by adding an up-sampling path, and extended DenseNets to FCN to reduce the impact of large computation and parameters, resulting from high-resolution characteristic patterns multiplied by a large number of input filters. However, these networks have difficulties in the segmentation of small and medium targets.

In view of this, researchers combined the detection and segmentation tasks to produce a series of networks. Mask RCNN (18) is the most advanced instance segmentation technology, with which, the pixel-level labels in each prediction box can be predicted, and multitask segmentation can be completed. However, its pooling layer using "ROI-align" reduces the resolution of input features, resulting in a loss of spatial detail of the image. In addition, it will increase computer overheads and network parameters. BlitzNet (19) used a full convolutional network to detect and segment the target and promoted model learning through weight sharing between the two tasks. But the model did not show the advantages of segmenting tasks. Based

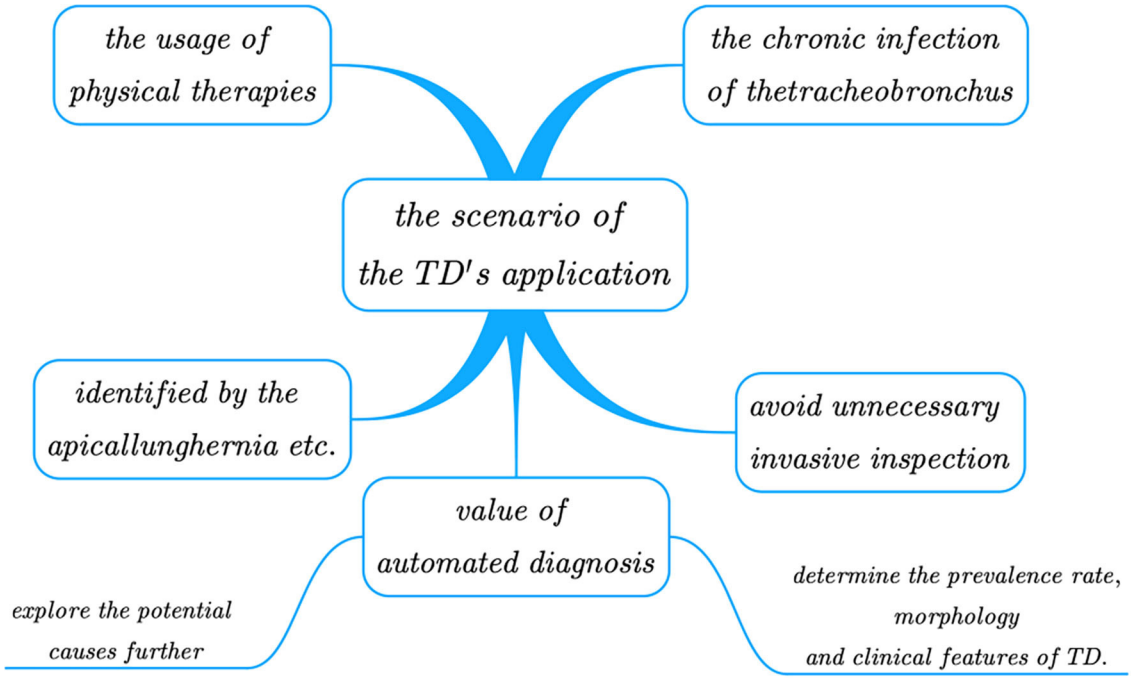


FIGURE 1 | The scenarios of the TD's application.

on the network mentioned above, Zhang et al. (20) introduced an ROI convolution directly in the location network area. This convolution module operates on all ROI during calculation, and it is applied to acetabulum segmentation from ultrasonic images. In addition, it helps solve the segmentation problem of small targets by experimentally demonstrating the introduction of localization units. Bak et al. (21) used the deep learning model to analyze the LDCT images of 131 patients, and calculated related indicators such as the emphysema index and mean lung density. Deep learning and statistical methods were also combined to detect emphysema quantitatively. Jin et al. (22) proposed a two-stage deep learning structure to improve the accuracy of emphysema quantification. Srinivasu et al. (23) proposed an algorithm based on probabilistic deep Q network for real-time robotic surgery, showing excellent performance in terms of the learning rate. Lin et al. (24) developed a deep learning model for COVID-19 identification and compared the CT images of TD and the coronavirus. Deep learning methods are also widely used in face recognition (25), breast cancer detection (26), and cervical cancer classification (27).

Research Status of Cloud Technology

With the development of cloud technology and blockchain technology, Ghayvat et al. (28) combined the blockchain (BC)-based confidentiality-privacy (CP) preserving scheme, which improved the accuracy by 12% compared with the traditional algorithm, verifying that it is a good solution for big data storage and privacy protection. At the same time, they also designed a COUNTERACT system that can calculate the incubation period of new crown objects (29) and an environmentally assisted living system (30) that well combines the latest sensor fusion technology. These designs will be the trend of future applications.

The Deficiency of Existing Research

Although improvements have been made in the performances of various methods based on deep learning, there is still room for breakthroughs in some areas. The motivations for the proposed methods can be summarized in the following two aspects:

First, the existing TD identification task relies mostly on the data collected by physicians in the diagnosis process, most of which needs to be visually observed by professionals and written into the diagnosis report. Since TD lesion sites are not obvious, identification and diagnosis by manual means are prone to fatigue, misdiagnosis, and missed diagnoses. In existing models of lung diseases based on deep learning, most of the focuses are on cancerous sites (31, 32), pulmonary nodules (33–35), etc. Many studies have ignored the effect of paramediastinal air cysts on the subsequent disease.

Secondly, the existing methods are not robust, but in TD semantic segmentation, the identifiable lesion covers only a small area with low pixels, and edge restoration is needed. Moreover, the TD's position in CT image needs to be identified, which can hardly be achieved by applying existing methods to these datasets. The advancement and the limitations of the current study can be concluded in **Table 1**.

The Novelties and Contributions of the Proposed Work

To address the above limitations, we propose a novel structure to classify every pixel of TD through a dual-channel attention-based

TABLE 1 | The conclusion of advancement and the limitations of the current study.

Advancement	Limitations
TD Case reports	Not robust
Machine learning, with deep learning focused on other areas of interest on regions	Few scholars have automated the diagnosis of TD

segmentation network. The main features of the proposed method are as follows:

1. Unlike the traditional bounding box-based target detection, or the entire image-based classification task, we performed the semantic segmentation on the TD lesions identification by classifying each pixel point by pixel. This method helps to better utilize the original data to present a clearer image, which is convenient for subsequent diagnosis. This proposed method is the first semantic segmentation network for TD diagnosis.
2. The proposed method saves the traditional intermediate steps (feature extraction) by taking the end-to-end loss function as the features of the generated evaluation and the GT confirmation. It can directly and automatically learn the overall lesion characteristics instead of manual identifications that undermine the generalization performance.
3. Given the outstanding performance of the attention mechanism on semantic segmentation tasks, we proposed a novel dual-channel attention TD segmentation network with position coding and contour coding modules, which has better accuracy and robustness. We compared the proposed model with the traditional various semantic segmentation models based on deep learning. The MA and MIOU of our method can be as high as 0.92 and 0.87, respectively. On this basis, we also carried out corresponding ablation experiments to verify the advantages of the added modules.

Paper Organization

The rest of the paper is organized as follows: first, in the "Experimental Methods and Materials" section, we introduced the data sources and preprocessing methods; also, the proposed model is elaborated in this section. As for the "Results and discussion" section, a series of comparative and ablation experiments are described. Finally, we summarized this work and pointed out the directions for future research in the conclusion section.

EXPERIMENTAL METHODS AND MATERIALS

Experimental Data and Preprocessing

The cases included in this study were patients who underwent routine chest CT scanning using Philips Brilliance 16- and 64-slice spiral CT machine in the Department of Medical Imaging, Fengxian People's Hospital, Jiangsu Province, from October 2020 to January 2021. The patients were scanned by CT from the superior border of the thyroid cartilage to the bottom of the lung. The field of vision was 350 mm, 120 kV, 250 mA, and the collimator width was 0.75 mm. The original data of each period were reconstructed with 1 mm layer thickness and 0.15 mm layer spacing, and postprocessing techniques such as multiplane recombination (MPR), maximum density projection, and volume reconstruction were applied at the workstation. The ground truth we annotated based on the original image was all annotated and verified by experienced radiologists. The image size is 512*512, and all data used for training, testing, and verification are desensitized. **Figure 2** shows a set of examples.

In the collected pathological images, the lesion area is highly obscure, where light environment or photographing can easily lead to background pollution, so it is challenging to segment the lesion area accurately. Therefore, we mainly used two methods to preprocess the data. Firstly, the data of HSV hue transformation were enhanced, which can avoid the influence of light on the image, making the network more robust and the color space clearer compared with the method of RGB color space. HSV can be described as follows:

$$\text{Output} = \text{HSV}(\text{x}_{\text{i=r,g,b}} \cdot \alpha_{\text{i=h,v,s}}) \tag{1}$$

where, x is the single channel of input image ($i = r, g, bi = r, g, b$); a ($j = h, s, vj = h, s, v$) is the disturbance factor; HSV is the augmented model of color space. In our experiment, $h = 0.015, s = 0.7, v = 0.4h = 0.015, s = 0.7, v = 0.4$.

Secondly, we adopted the nonlinear signal processing technique based on ranking statistical theory to suppress noise for effective preprocessing. Specifically, in the small lesion area, it is more prominent with the effect of salt and pepper noise. Therefore, we created a sliding window near the lesion area, which traverses all pixels of the area, accumulates all pixel values, and takes the median one as the enhanced pixel. It can effectively protect the edge information of the image while suppressing noise points. The equation can be expressed as:

$$\text{h}(\alpha, \beta) = \text{med}\left\{f(\alpha - m, \beta - n), m, n \in W\right\} \tag{2}$$

where $f(\alpha, \beta)$ and $h(\alpha, \beta)$ denote the original image and the processed image, respectively, and W is the two-dimensional input template. *med* represents the convolutional implementation of the sliding window. All the original images used for network design were 512 × 512 pixel in size, with a horizontal and vertical resolution of 96 dpi. The image size was kept the same to accommodate the CNN network training.

Network Structure

Overall Structure

This part mainly introduced the network based on attentional tracheobronchial diverticulum segmentation, as shown in **Figure 3**. Firstly, the resnet101 network was employed as the backbone to extract the deep and shallow features of the image. In addition, considering the smallness of the lesion and its complex background, a two-channel network was designed to achieve the segmentation task with both rough and fine-grained results, which consisted of two parts—one was location feature extraction and encoding module, which extracts multiscale spatial features of the object to be segmented and locating the target area in the

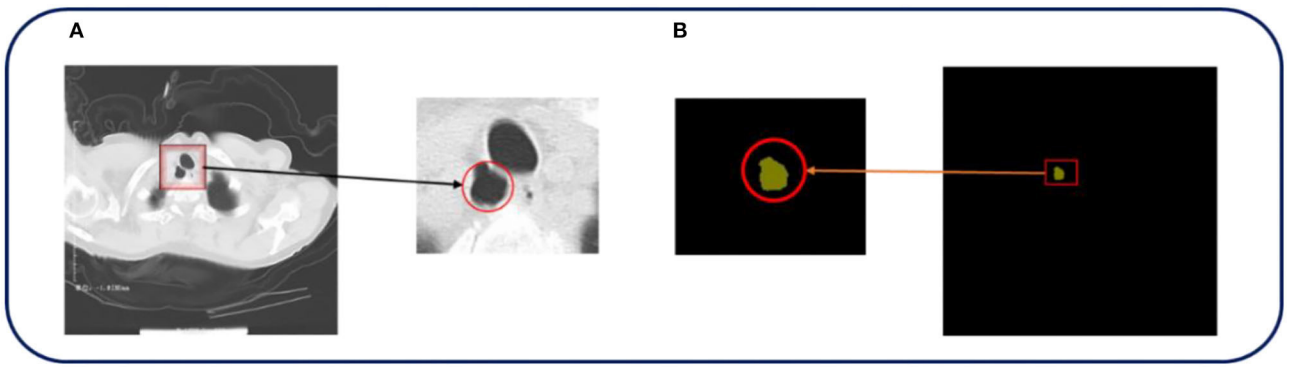


FIGURE 2 | The original CT image and the corresponding ground truth are shown in **Figure 1**. **(A,B)** In the above pictures, red represents TD, which is the part that we are interested in.

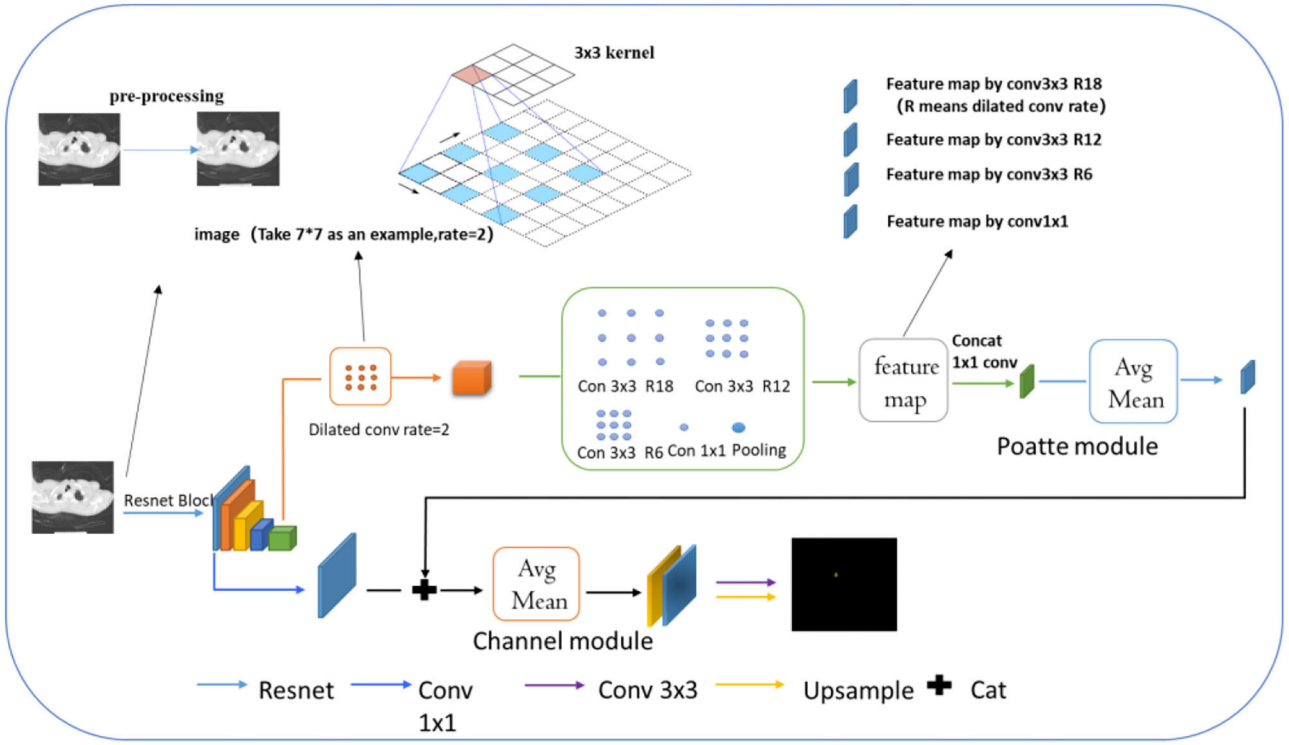


FIGURE 3 | Attention-based tracheobronchial diverticulum segmentation network.

image, and the other was the contour encoding module, which was in charge of fusing and upsampling (mapping) the output of the initial Conv-2 feature of Resnet101 and the location encoding module, so as to further enhance abilities of details recovery of contour edges of the image.

Further explanation about the model is made in the following. Before extracting features by Resnet, we used the preprocessing model mentioned above to process the TD image. Dilated convolution is a common one whose working mode is explained in the image by arrows. Compared with conventional convolutions, its receptive field offers more lesion information. Then, different feature maps by various types of dilated convolution can be received, which are processed by concat in the next step.

Location Encoding Module (Poatte Module)

Using Resnet101, we extracted features of deep and shallow layers. In the location encoding module, the target area should be first located in the image, around the left hand and close to the surrounding trachea and the right upper lung tip. Then we employed the deep layer to extract the trachea and right upper lung tip features and further identified the esophagus

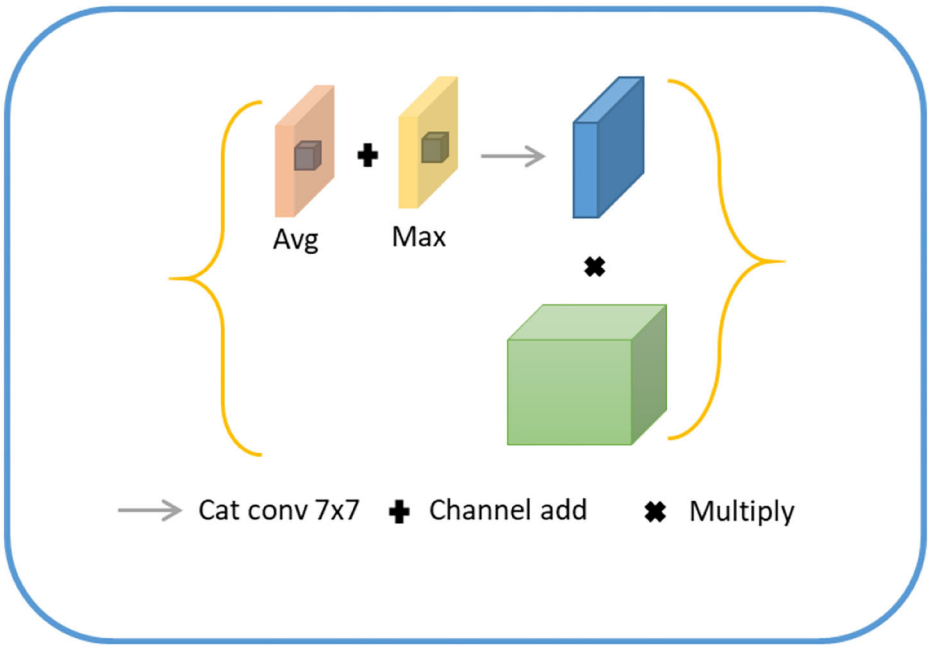


FIGURE 4 | Schematic diagram of location coding module.

area. However, in this layer, smaller targets tended to disappear after multiple convolution pooling. For that, the dilation (ASPP) convolution (36) (expansion rate of 2) was used to increase the receptive field of the image and decrease the probability of the disappearance of small targets in the deep feature map. In the case of two-dimensional signals, for each position *i* and *a* convolution filter *w* on the output feature map *y*, they were applied by atrous convolution on the input feature map *x*, as follows.

$$y[i] = \sum_{k} x[i + r \cdot k]\, w[k] \tag{3}$$

In the above function, the atrous rate *r* determined the stride, with which we sample the input signal, and the output y of ASPP through the position module was:

$$Ms(F) = \sigma\left(f\left(\left[AvgPool(F); MaxPool(F)\right]\right)\right) \tag{4}$$

$$Ms(F) = \sigma\left(f1 \times 1\left(\left[Fs\ avg; Fs\ max\right]\right)\right) \tag{5}$$

In the formula, f1 × 1 was the convolution kernel of 1 × 1, and σ was the sigmoid function.

Besides, the chosen single deep scale was conducive to large target detection but poor for the small one. On this single scale, what 3 × 3 convolution and 1 × 1 convolution can capture was only local region features rather than the global ones. Therefore, the feature map obtained by the upper expansion convolution was conducted by different expansion rates. With the increase of convolution layers, the convolution kernel degenerated into 1 × 1 when the expansion rate of convolution was large, which could not capture global ones. To solve the problem caused, we first added image-level features, specifically, the average pooling of each channel of the feature map, and then the original scale resolution was sampled up. Secondly, we set the expansion rates of these convolutions to 6, 12, 18, and ultimately fused the average pooling features with the features of different expansion rates. The fused features possessed characteristics of global context so that the target at different scales can be roughly located. Meanwhile, because the ultimate goal was to find small lesion areas around the large target esophagus, fused features were put into the location attention module, as shown in **Figure 4**. The module first pooled the input features into global maximum and global average pooling and fused them into 7 × 7 convolutions, which were positionally expressed, and this generally made the target location more salient. Finally, the information of location was multiplied by the characteristics of the previous layer to further identify the small target position.

Contour Coding Module (Channel Module)

For lesion being a small object, the spatial information of the target was usually lost after deep convolution so that shallow features were employed to restore the contour information of the target. Specifically, we first conducted a 1 × 1 convolution of shallow features, fused with the input of the position-coding module to endow them with the contour information of shallow features and position of deep features. Next, the fused feature was processed through the attention module of channel fusion, as shown in **Figure 5**, where two 1 × 1 × C feature maps were obtained by planar and maximum pooling of the fused features in terms of the long and wide directions. The final channel feature was strengthened by a shared convolution layer and pixel-level sum operation, making the initial feature profile the target area.

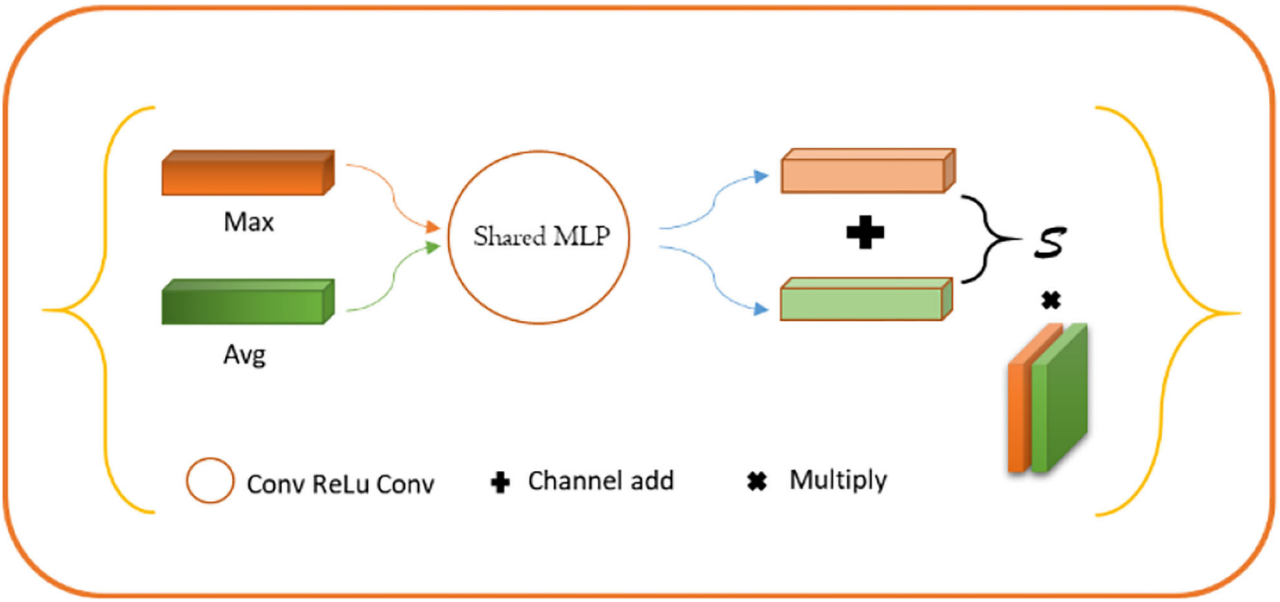


FIGURE 5 | Schematic diagram of contour coding module.

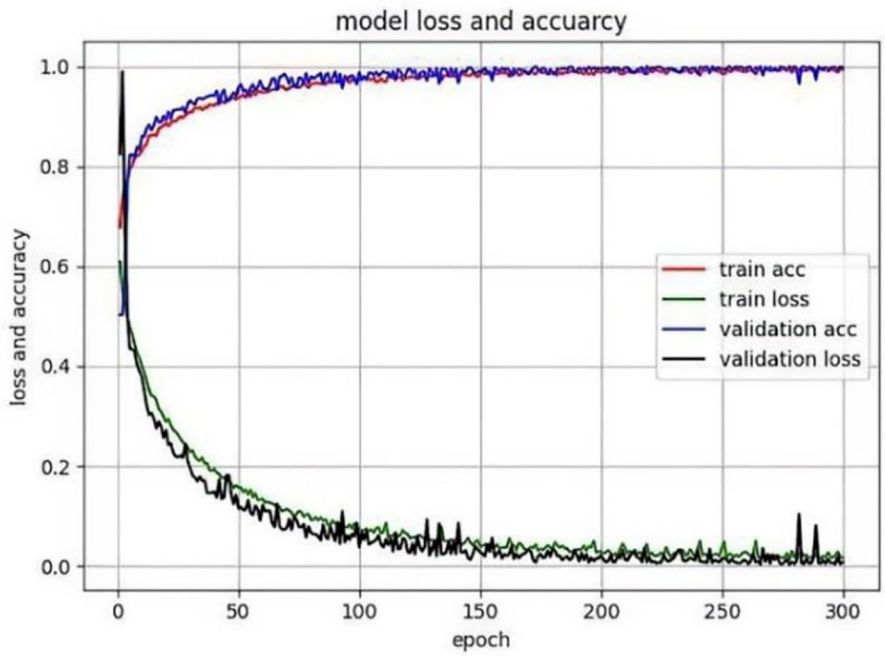


FIGURE 6 | Model training chart.

Finally, the feature map was recovered to the original scale size by convolution and upsampling.

Experiment Settings

The method was implemented based on the Pytorch library and ubantu environment, with the learning rate being 0.0001, the momentum and weight decay coefficients 0.9 and 0.0001, respectively. The model was trained by GTX 2080 Ti, Batchsize = 4, and the training epoch was 300. When the epoch was 0 to 150, the model converged quickly, improving accuracy. But it was relatively stable and converged moderately with slow-moving accuracy between 150 and 300, as shown in **Figure 6**. In addition,

the TD lesion dataset was obtained from the hospital, and 218 images (190 images in training set, 10 images in validation set, and 18 images in test set) and 218 lesion area markers were constructed under the guidance of pathologists.

Based on this, the pretraining model with fine adjustment is applicable in a resource-constrained environment. The resources consumed are only the computational ones that are returned in the following gradient. Moreover, the following multiscale module of feature extraction and feature attention do not require a great number of computing resources. In the multiscale feature extraction module, we used dilated convolution to expand the receptive field and decrease the representation of feature maps, and even the demand for resources. We repeatedly utilized 1*1 convolution in the module to compress features to integrate representation of features and lower demands for resources. Based on the above two points, the model can be trained and tested in a resource-constrained environment.

Evaluation Indexes

We adopted two most essential evaluation indexes of semantic segmentation, namely, mean pixel-level accuracy (MPA) and mean intersection over union (MIoU), which were as follows :

$$\mathrm{MPA} = \frac{1}{\mathrm{T}}\sum_{\mathrm{T}=1}^{\mathrm{T}}\left[\frac{1}{\mathrm{C}+1}\sum_{\mathrm{m}=0}^{\mathrm{C}}\frac{\mathrm{p}^{(\mathrm{mm})}}{\sum_{\mathrm{n}=0}^{\mathrm{C}}\mathrm{p}^{(\mathrm{nn})}}\right] \tag{6}$$

$$\mathrm{MIoU} = \frac{1}{\mathrm{T}}\sum_{\mathrm{T}=1}^{\mathrm{T}}\left[\frac{1}{\mathrm{C}+1}\sum_{\mathrm{m}=0}^{\mathrm{C}}\frac{\mathrm{p}^{(\mathrm{mm})}}{\sum_{\mathrm{n}=0}^{\mathrm{C}}\mathrm{p}^{(\mathrm{mn})}+\sum_{\mathrm{n}=0}^{\mathrm{C}}\mathrm{p}^{(\mathrm{nm})}-\mathrm{p}^{(\mathrm{mm})}}\right] \tag{7}$$

In the above functions, $p^{(mn)}$ denotes the true value of m, predicting the correct number of TD lesions. $C + 1$ was the number of categories (including empty classes). T was the number of CT images of the tracheobronchial diverticulum, which was 218 in this study. *MPA* was the average value of the correct classification of calculating pixel categories, *MIoU* was generally calculated based on the categories, that is, the *IoU* of each category was accumulated after the calculation, and then their average was the holistic evaluation.

RESULTS AND DISCUSSION

Comparison With Advanced Semantic Segmentation Algorithms

In this part, we selected the classic and novel semantic segmentation networks, including FCN, DENSE, DANet, PSPNet, and DoubleUnet. Among these networks, FCN (37) was the first network that applied CNN structure to semantic segmentation tasks and altered many classification networks into fully convolutional networks. Then, by introducing Atrous convolution, DENSE solved the coding problem of multiscale information and obtained the spatial pyramid structure suitable for semantic segmentation. DANet (38) proposed a dual attention network for segmentation, which could introduce spatial and positional attention to adjust the local semantic information in the adaptive integrated image. PSPNet (39) proposed a pyramid network that could extract context semantic information, improving the segmentation performance in complex scenes. DoubleUnet (40) employed two stacked U-Net (one pretrained coding network was used to capture more abundant semantic information, and the other was set at the bottom to extract information) to improve the segmentation performance of multiple sets of medical images.

TABLE 2 | Results of common semantic segmentation methods in identifying tracheobronchial diverticula.

Method	Backbone	MPA	MIoU
FCN32S	VGG16	0.77	0.50
FCN16S	VGG16	0.74	0.49
FCN8S	VGG16	0.75	0.49
FCN	ResNet50	0.78	0.59
FCN	ResNet50	0.81	0.53
DENSE	DENASPP121	0.87	0.64
DENSE	DENASPP161	0.89	0.78
DENSE	DENASPP169	0.78	0.62
DENSE	DENASPP201	0.78	0.63
DANet	ResNet50	0.77	0.55
DANet	ResNet101	0.77	0.58
DANet	ResNet152	0.88	0.63
PSPNet	ResNet50	0.60	0.34
PSPNet	ResNet101	0.63	0.34
PSPNet	ResNet152	0.63	0.34
DoubleUnet	ResNet50	0.78	0.53
Ours	ResNet101	**0.92**	**0.87**

The meaning of the bold values is that the network we designed in this article (We call it ours for short).

To verify the effectiveness of our dual-channel network, MPA and MIOU were adopted as evaluation indexes to compare (SOTA) FCN32S, FCN16S, FCN8S, FCN, DENSE, DANet, PSPNet, DoubleUnet network, and build different algorithms. We selected VGG-16 (41) and ResNet-50 (42) as feature extraction networks, respectively. The former obtained more semantic information by deepening the network depth. The latter solved the problem of gradient explosion and disappearance by introducing residual structure and enhancing the generalization performance of the network with wide applications in classification and segmentation tasks in many scenarios. We conducted the experiment combining the above backbones with the segmentation network, and the results are shown in **Table 2**. It can be seen from **Table 1** that the MIOU value of FCN network, after establishing VGG16 and ResNet50, was about 0.4–0.5, the figure after DENASPP121,161,169,201 as about 0.6–0.7, the one after ResNet series network was about 0.5–0.6, the figure after ResNet series network was approximately 0.3, and the accuracy of DoubelUnet network was about 0.53. However, the MIOU accuracy of the proposed network was 0.87, much higher than the currently popular algorithms, mainly due to the effective extraction of spatial features and the efficient use of boundary information in the network.

In addition, Resnet50, Mobilnetv2 (43), Drn16 (44), and Xcepotionv2 (45) were also selected as backbones for the dual-channel network. As for their results, the highest MIOU was 0.87 and the lowest was 0.77, both lower than the proposed algorithm, showing the robustness and efficiency of the proposed network. The comparison is shown in **Table 3**.

First of all, we compared the indexes of different backbones, and their good performance proved the universality and validity of the segmentation of modules. ASPP module was proposed for the multiscale problem of objects. Before the convolution, it was resampled at a given feature layer with multiple sampling rates. We used multiple paralleled dilated convolution layers with different sampling rates to expand the receptive field and enhance the expression of features. In the model, the convolution with a convolution kernel of 1 was used repeatedly to compress, integrate, and increase nonlinearity. In view of the obscure feature points of the object, we adopted the global average pooling to enhance the response in the channel further, making the feature more prominent and conducive to the segmentation. Channel attention is to strengthen the learning of different fields of vision features. Furthermore, spatial attention is to enhance the learning of spatial expression features of objects and the response of feature points. Based on the above points, the model we designed strengthens feature learning, thereby improving segmentation accuracy.

TABLE 3 | Comparing the results of our model with different backbones for identifying tracheobronchial diverticula.

Method	Backbone	MPA	MIoU
Ours	Resnet50	0.90	0.87
Ours	Mobilenetv2	0.90	0.83
Ours	Drn16	0.92	0.87
Ours	Xceptionv2	0.84	0.77

For a more direct exhibition of the algorithm, we selected algorithms used for the above comparison and the TD lesion segmentation map (**Figure 7**) by the improved algorithm, which was compared with ground truth.

In the experiment, we found that the recognition results of popular algorithms (such as DoubleUnet) were overfitting, causing the incomplete identification of the lesion area for its smallness and distraction of organs around it. However, the network was designed by setting a position-coding module and a contour-coding module, introducing positional and spatial attention to focus on extracting the semantic information of the lesion area, which made its segmentation effect better than the existing algorithms.

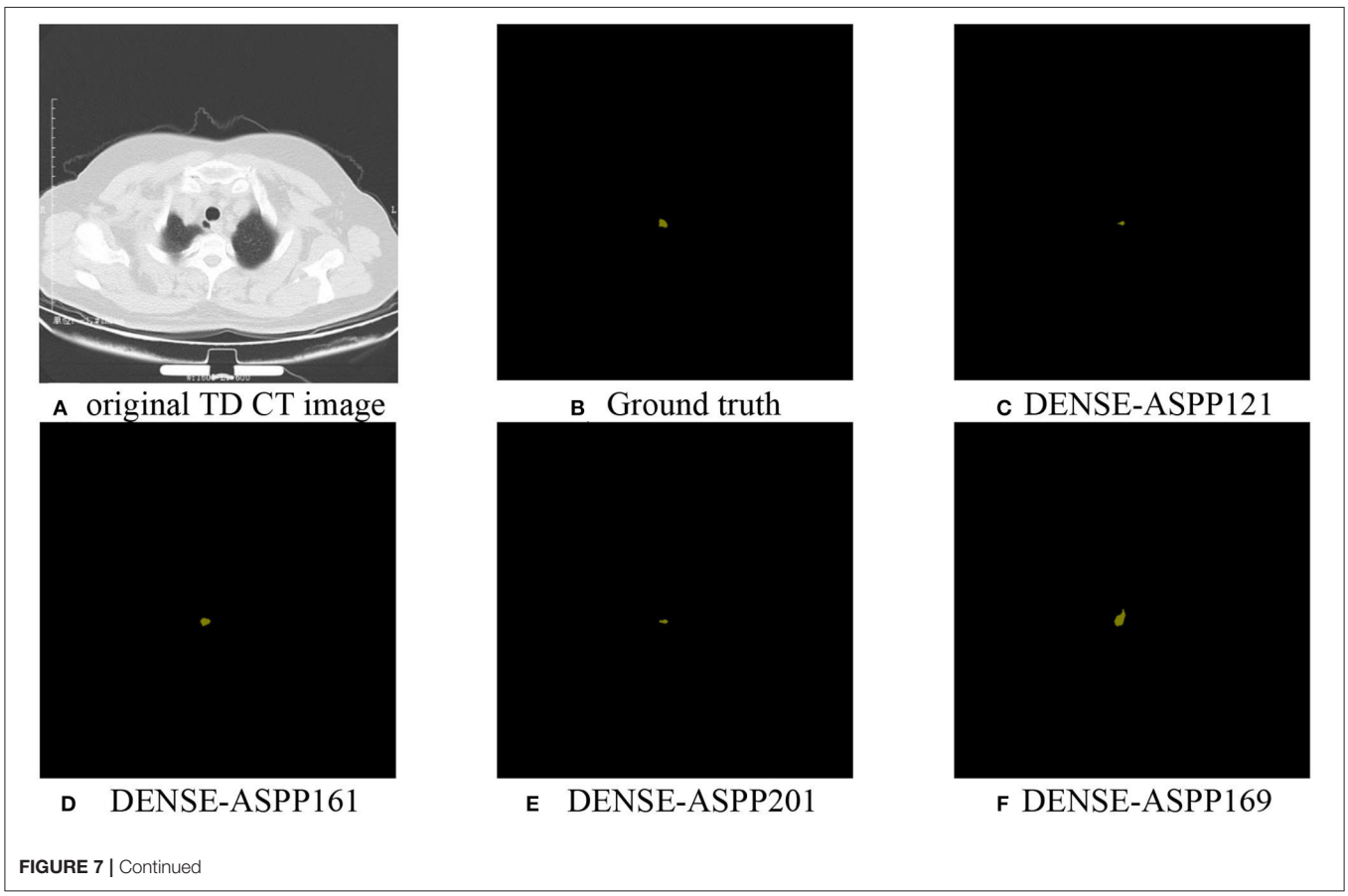


FIGURE 7 | Continued

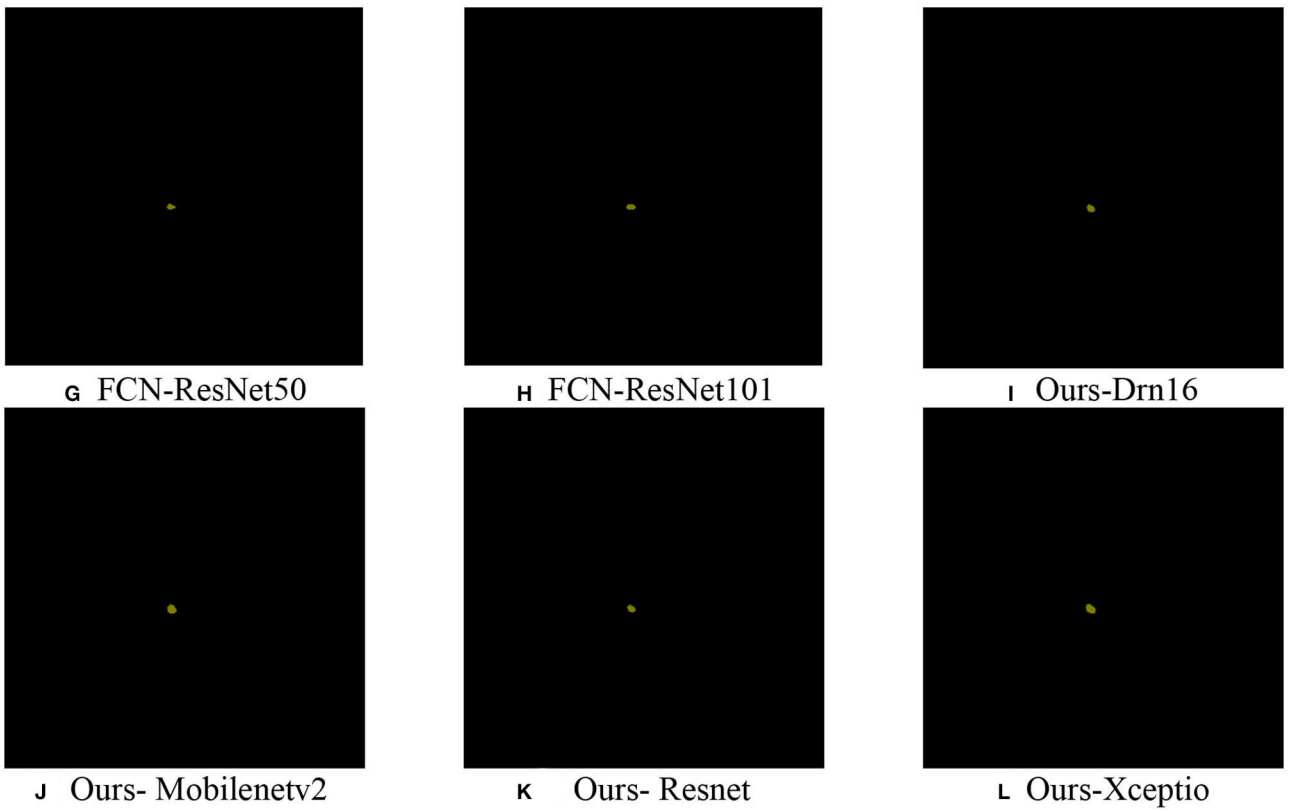


FIGURE 7 | TD lesion segmentation diagram: **(A)** the original CT data; **(B)** ground truth; **(C–H)** partially selected comparison algorithm; **(I–L)** predicted results using Drn16, Mobilenetv2, Resnet, and Xception as the encoder extraction network, respectively.

TABLE 4 | Comparison of the results of our model with different backbones for identifying tracheobronchial diverticula.

Atrous convolution	PCM	CAM	MIoU
×	×	×	0.68
√	×	×	0.79
×	√	√	0.82
√	√	√	0.87

Ablation Experiment

Atrous convolution and positional attention module, and spatial attention module (PCM) were added to the network. To verify the effectiveness of the two modules, we performed the following ablation experiments in **Table 4**. When there was no module in the network, the evaluation index MIOU was 0.68. When deep and separable convolution was added, it was 0.79, and the accuracy value was higher than that of the original network 0.11. The reason was that it selected convolutions with different expansion rates when extracting features at high scales to increase the receptive field, making features more holistic and avoiding the loss of target information. When positional and spatial attention modules (PCM) were added, the MIOU accuracy was 0.82 without dilated convolution, higher than 0.14 of the original network. The reason behind it was that the two attentional modules made the target position prominent. Combining the deep and separable convolution (convolution with different expansion rates, Atrous convolution), positional attention module, and spatial attention module (PCM) with the network, the accuracy was 0.87, higher than the original accuracy of 0.68. The ablation experiment further verified the effectiveness of the modules.

Furthermore, combined with the visualized results of the above segmentation, it can be clearly found that the proposed algorithm can effectively locate and segment TD lesions. Meanwhile, the addition of attentional modules improved performance, especially reducing the false detection of small objects and providing richer positional, spatial, and background information than the ordinary network.

In addition, we drew a box diagram (**Figure 8**) to demonstrate the overall and critical features of the data. There were Danet, Dence, Fcn, PSPNet, and the MIoU indicators of segmentation network under different backbones, and all model indexes were drawn into a box diagram.

It can be seen from the five models that the median accuracy of PSPNet algorithm was only 30%, but the lower

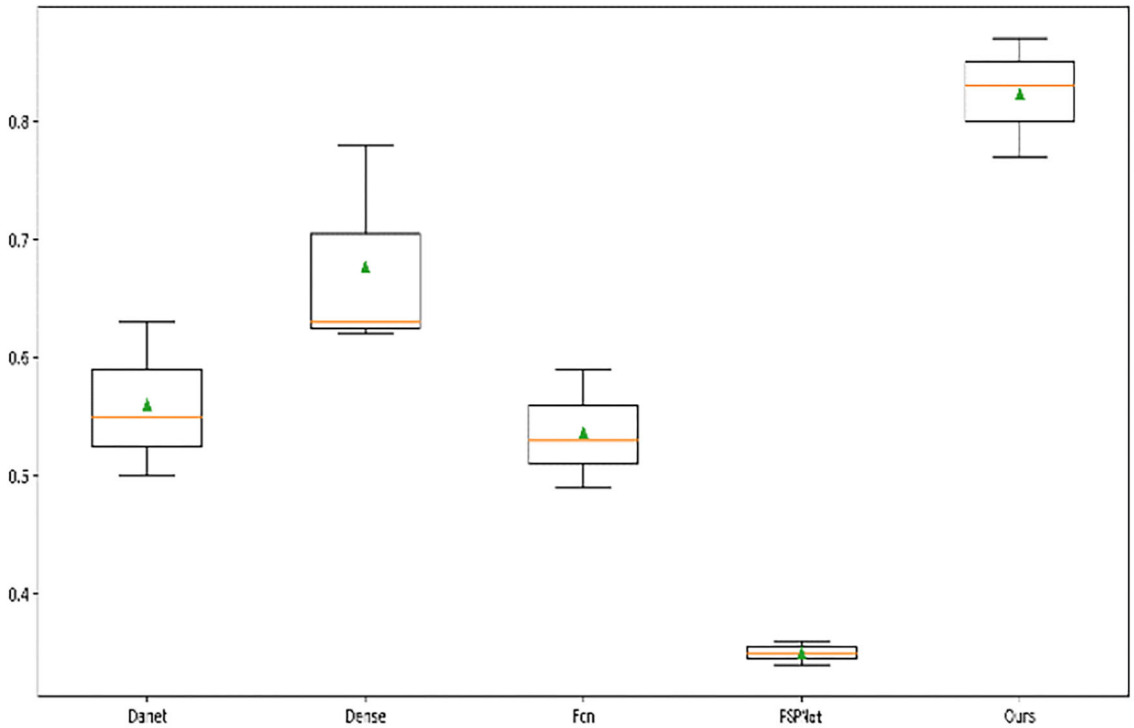


FIGURE 8 | Box diagram of different model indexes.

quartile accuracy of Danet, Dense, and FCN were higher than 50%. The median of the algorithm was higher than 80%, displaying that our attention-based strategy had better performance in most samples and showed the robustness of the algorithm. Although other algorithms have achieved excellent performance on many datasets, the overall recognition effect of TD lesions is not satisfying. It is believed that the reason is that TD lesion is a small target and easily confused with other surrounding organs, especially the esophagus area. Even many experienced doctors often confuse or fail to identify the lesion, which results in obstacles to later diagnosis of diseases.

CONCLUSION AND FUTURE WORK

This work presented a new method of TD lesion recognition based on CT image semantic segmentation. Most of the previous studies focused on case reports, rather than using the deep learning method for segmentation to reduce the workload of doctors. In terms of that, we conducted pixel-level segmentation for the lesion area of CT images (that is, semantic segmentation) instead of traditional classification and target detection tasks. Inspired by the great success of attentional mechanism in image segmentation, we adopted the dual-channel attention model to automatically identify TD lesions and different pretrained feature extraction networks, including VGG16, DENASPP121, and ResNet50 as the backbone of the encoder, to improve the prediction performance of the network. Also, the position-coding and contour-coding module were redesigned with the integration of spatial and positional attention into the network, improving the identification of lesion position with coarse and fine-grained information. It has been proved that the segmentation network with Resnet, Mobilnetv2, Drn16, and Xcepotionv2 as the backbone of the encoder and the proposed coding module can effectively restore the details of the lesion area, superior to other popular semantic segmentation models, and novel algorithms based on U-net deformation. In other words, the value of MioU was 0.87, fully explaining that the semantic segmentation method is conducive to identifying TD lesions.

However, it should be noted that the research still has some limitations. First of all, because the above CT image resolution was low even with occlusions somewhere, it will be helpful to improve the network's performance once they are improved. Furthermore, more complex CT data should be introduced to enhance the robustness of the model. Ultimately, other ignored diseases like TD deserve further study so that deep learning can be applied to disease diagnosis and be integrated into a network with stronger generalization and migration ability. Once similar lesions with fixed positions and small targets are met, we can segment the network to obtain the needed results. Moreover, we will design an application to assist in the diagnosis of TD, with the aim of improving the diagnosis efficiency of medical practitioners.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual for the publication of any potentially identifiable images or data included in this article. The studies involving human participants were reviewed and approved by the Fengxian People's Hospital. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CD: resources and project administration. SG: overall experimental idea design and article editing. MZ: writing—original draft preparation and verification. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Medical Research Project of Jiangsu Provincial Health Department (YG201419).

ACKNOWLEDGMENTS

We thank CD and Fengxian People's Hospital for providing the original medical dataset for our work.

REFERENCES

1. Toscano L, Terra D, Salisbury S, Arechavaleta N. Surgical resection of tracheal diverticulum with haemoptysis as unusual presentation. *Case Rep Surg.* (2019) 2019:3828197. doi: 10.1155/2019/3828197
2. Buterbaugh J, Erly W. Paratracheal air cysts: a common finding on routine CT examinations of the cervical spine and neck that may mimic pneumomediastinum in patients with traumatic injuries. *Am J Neuroradiol.* (2008) 29:1218–21. doi: 10.3174/ajnr.A1058
3. Polat AV, Elmali M, Aydin R, Ozbay A, Celenk C, Murat N. Paratracheal air cysts: Prevalence and correlation with lung diseases using multi-detector CT. *J Med Imaging Radiat Oncol.* (2014) 58:144–8. doi: 10.1111/1754-9485.12095
4. Boyaci N, Dokumaci DS, Karakas E, Yalcin F, Kurnaz AGO. Paratracheal air cysts: prevalence and relevance to pulmonary emphysema and bronchiectasis using thoracic multidetector CT. *Diagn Int Radiol.* (2015) 21:42. doi: 10.5152/dir.2014.14152
5. Bae HJ, Kang EY, Yong HS, Kim YK, Woo OH, Oh YW, et al. Paratracheal air cysts on thoracic multidetector CT: incidence, morphological characteristics and relevance to pulmonary emphysema. *Br J Radiol.* (2013) 86:20120218. doi: 10.1259/bjr.20120218
6. Mineshita M, Tajima O, Kondo T. Paratracheal air cysts in middle-aged Japanese men. *J Bronchology Interv Pulmonol.* (2006) 13:6–8. doi: 10.1097/01.laboratory.0000199580.68036.eb
7. Pace M, Dapoto A, Surace A, Di Giacomo A, Morzenti C, Costantini E, et al. Tracheal diverticula: a retrospective analysis of patients referred for thoracic CT. *Medicine.* (2018) 97:e12544. doi: 10.1097/MD.0000000000012544
8. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput.* (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
9. Dhanamjayulu C, Nizhal UN, Maddikunta PK, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021). doi: 10.1049/ipr2.12222
10. Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, et al. IoT for Smart Cities: machine learning approaches in smart healthcare—a review. *Fut Internet.* (2021) 13:218. doi: 10.3390/fi13080218
11. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HM, et al. Hep-Pred: hepatitis C staging prediction using fine Gaussian SVM. *Comput Mater Cont.* (2021) 69:191–203. doi: 10.32604/cmc.2021.015436
12. Shah A, Ahirrao S, Pandya S, Kotecha K, Rathod S. Smart cardiac framework for an early detection of cardiac arrest condition and risk. *Front Public Health.* (2021) 9:762303. doi: 10.3389/fpubh.2021.762303
13. Hasan MK, Ghazal TM, Alkhalifah A, Bakar KA, Omidvar A, Nafi NS, et al. Fischer linear discrimination and quadratic discrimination analysis-based data mining technique for internet of things framework for healthcare. *Front Public Health.* (2021) 9:737149. doi: 10.3389/fpubh.2021.737149
14. Saxe A, Nelli S, Summerfield C. If deep learning is the answer, what is the question? *Nat Rev Neurosci.* (2021) 22:55–67. doi: 10.1038/s41583-020-00395-8
15. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing And Computer-Assisted Intervention.* (2015). p. 234–41
16. Shrivastava R, Kumar P, Tripathi S, Tiwari V, Rajput DS, Gadekallu TR, et al. novel grid and place neuron's computational modeling to learn spatial semantics of an environment. *Appl Sci.* (2020) 10:5147. doi: 10.3390/app10155147
17. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* (2017). p. 11–9
18. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE International Conference on Computer Vision.* (2017). p. 2961–2969
19. Dvornik N, Shmelkov K, Mairal J, Schmid C. Blitznet: a real-time deep network for scene understanding. In: *Proceedings of the IEEE International Conference on Computer Vision.* (2017). p. 4154–62.
20. Zhang Z, Tang M, Cobzas D, Zonoobi D, Jagersand M, Jaremko JL. End-to-end detection-segmentation network with ROI convolution. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* (2018). p. 1509–12
21. Bak SH, Kim JH, Jin H, Kwon SO, Kim B, Cha YK, et al. Emphysema quantification using low-dose computed tomography with deep learning–based kernel conversion comparison. *Eur Radiol.* (2020) 30:6779–87. doi: 10.1007/s00330-020-07020-3
22. Jin H, Heo C, Kim JH. Deep learning-enabled accurate normalization of reconstruction kernel effects on emphysema quantification in low-dose CT. *Phys Med Biol.* (2019) 64:135010. doi: 10.1088/1361-6560/ab28a1/meta
23. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic Deep Q Network for real-time path planning in censorious robotic procedures using force sensors. *J Real-Time Image Process.* (2021) 18:1773–85. doi: 10.1007/s11554-021-01122-x
24. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology.* (2020) 296:E65–71. doi: 10.1148/radiol.2020200905
25. Khalifa OO, Omar AA, Ahmed MZ, Saeed RA, Hashim AH, Esgiar AN. An automatic facial age proression estimation system. In: *2021 International Congress of Advanced Technology and Engineering (ICOTEN).* IEEE (2021). p. 1–6.
26. Siddiqui SY, Haider A, Ghazal TM, Khan MA, Naseer I, Abbas S, et al. IoMT Cloud-based intelligent prediction of breast cancer stages empowered with deep learning. *IEEE Access.* (2021) 9:146478–91. doi: 10.1109/ACCESS.2021.3123472

27. Khamparia A, Gupta D, de Albuquerque VH, Sangaiah AK, Jhaveri RH. Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning. *J Supercomput.* (2020) 76:8590–608. doi: 10.1007/s11227-020-03159-4
28. Ghayvat H, Pandya SN, Bhattacharya P, Zuhair M, Rashid M, Hakak S, et al. CP-BDHCA: blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications. *IEEE J Biomed Health Inform.* (2021) 2021:1. doi: 10.1109/JBHI.2021.3097237
29. Ghayvat H, Awais M, Gope P, Pandya S, Majumdar S. Recognizing suspect and predicting the spread of contagion based on mobile phone location data (counteract): a system of identifying covid-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence. *Sustain Cities Soc.* (2021) 69:102798. doi: 10.1016/j.scs.2021.102798
30. Ghayvat H, Awais M, Pandya S, Ren H, Akbarzadeh S, Chandra Mukhopadhyay S, et al. Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection. *Sensors.* (2019) 19:766. doi: 10.3390/s19040766
31. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med.* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x
32. Shin H, Oh S, Hong S, Kang M, Kang D, Ji YG, et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano.* (2020) 14:5435–44. doi: 10.1021/acsnano.9b09119
33. Massion PP, Antic S, Ather S, Arteta C, Brabec J, Chen H, et al. Assessing the accuracy of a deep learning method to risk stratify indeterminate pulmonary nodules. *Am J Respir Crit Care Med.* (2020) 202:241–9. doi: 10.1164/rccm.201903-0505OC
34. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology.* (2019) 290:218–28. doi: 10.1148/radiol.2018180237
35. Qi LL, Wu BT, Tang W, Zhou LN, Huang Y, Zhao SJ, et al. Long-term follow-up of persistent pulmonary pure ground-glass nodules with deep learning-assisted nodule segmentation. *Eur Radiol.* (2020) 30:744–55. doi: 10.1007/s00330-019-06344-z
36. Yang M, Yu K, Zhang C, Li Z, Yang K. Denseaspp for semantic segmentation in street scenes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2018). p. 3684–92.
37. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2015). p. 3431–40.
38. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H. Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* (2019). p. 3146–54.
39. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2017). p. 2881–90.
40. Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD. Doubleu-net: a deep convolutional neural network for medical image segmentation. In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS).* (2020). IEEE, p. 558–64.
41. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 (2014).
42. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 770–778
43. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861. (2017).
44. Dai B, Zhang Y, Lin D. Detecting visual relationships with deep relational networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition.* (2017). p. 3076–3086.
45. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2017). p. 1251–58.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Ding and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 06 January 2022
doi: 10.3389/fpubh.2021.813625

The Construction and Development of a Clinical Prediction Model to Assess Lymph Node Metastases in Osteosarcoma

Wenle Li[1,2†], Shengtao Dong[3†], Bing Wang[2], Haosheng Wang[4], Chan Xu[2], Kai Zhang[1,2], Wanying Li[2], Zhaohui Hu[5], Xiaoping Li[6], Qiang Liu[1*], Rilige Wu[7*] and Chengliang Yin[8*]*

[1] Department of Orthopedics, Xianyang Central Hospital, Xianyang, China, [2] Clinical Medical Research Center, Xianyang Central Hospital, Xianyang, China, [3] Department of Spine Surgery, Second Affiliated Hospital of Dalian Medical University, Dalian, China, [4] Department of Orthopaedics, The Second Hospital of Jilin University, Changchun, China, [5] Department of Spine Surgery, Liuzhou People's Hospital, Liuzhou, China, [6] Shulan International Medical College, Zhejiang Shuren University, Hangzhou, China, [7] College of Information and Electrical Engineering, China Agricultural University, Beijing, China, [8] Faculty of Medicine, Macau University of Science and Technology, Macau, Macao SAR, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Kun Liu,
Beijing Friendship Hospital, China
Zhidong Zhu,
Shanghai Outdo Biotech Co, China
Praveen Kumar,
VIT University, India

***Correspondence:**
Chengliang Yin
chengliangyin@163.com
Rilige Wu
wrlg1114@163.com
Qiang Liu
m13992079668@163.com
Xiaoping Li
li-xp@zjsru.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 12 November 2021
Accepted: 06 December 2021
Published: 06 January 2022

Citation:
Li W, Dong S, Wang B, Wang H, Xu C, Zhang K, Li W, Hu Z, Li X, Liu Q, Wu R and Yin C (2022) The Construction and Development of a Clinical Prediction Model to Assess Lymph Node Metastases in Osteosarcoma. Front. Public Health 9:813625. doi: 10.3389/fpubh.2021.813625

Background: This study aimed to construct a clinical prediction model for osteosarcoma patients to evaluate the influence factors for the occurrence of lymph node metastasis (LNM).

Methods: In our retrospective study, a total of 1,256 patients diagnosed with chondrosarcoma were enrolled from the SEER (Surveillance, Epidemiology, and End Results) database (training cohort, $n = 1{,}144$) and multicenter dataset (validation cohort, $n = 112$). Both the univariate and multivariable logistic regression analysis were performed to identify the potential risk factors of LNM in osteosarcoma patients. According to the results of multivariable logistic regression analysis, A nomogram were established and the predictive ability was assessed by calibration plots, receiver operating characteristics (ROCs) curve, and decision curve analysis (DCA). Moreover, Kaplan-Meier plot of overall survival (OS) was plot and a web calculator visualized the nomogram.

Results: Five independent risk factors [chemotherapy, surgery, lung metastases, lymphatic metastases (M-stage) and tumor size (T-stage)] were identified by multivariable logistic regression analysis. What's more, calibration plots displayed great power both in training and validation group. DCA presented great clinical utility. ROCs curve provided the predictive ability in the training cohort (AUC = 0.805) and the validation cohort (AUC = 0.808). Moreover, patients in LNN group had significantly better survival than that in LNP group both in training and validation group.

Conclusion: In this study, we constructed and developed a nomogram with risk factors, which performed well in predicting risk factors of LNM in osteosarcoma patients. It may give a guide for surgeons and oncologists to optimize individual treatment and make a better clinical decision.

Keywords: nomogram, SEER, osteosarcoma, lymph node metastases, multicenter

INTRODUCTION

Osteosarcoma is a common malignant bone tumor. The primary treatment consisting of neoadjuvant therapy, surgery and postoperative chemotherapy have resulted in the 5-year overall survival rate of ∼60% (1, 2). However, even with the treatment of surgery and chemotherapy, the prognosis for patients with metastatic osteosarcoma remains dismal (3, 4).The lung metastases, the primary target of metastasis in osteosarcoma, has five-year survival rates of ∼30% (5, 6). In extrapulmonary metastatic osteosarcoma, patients with lymph node metastases (LNM) have worse clinical outcomes, with five-year survival rates of 10% (7). However, only 3% of patients with osteosarcoma are diagnosed with LNM, leading to the lack of adequate clinical data for exploring osteosarcoma LNM (8).Therefore, a population-based study to assess the LNM in osteosarcoma is imminent.

Furthermore, disease forecasting is a vital part of the medical research (9–18). As a visual prediction tool, nomogram lists each variable separately and assigns a corresponding score for each status (19). Based on these considerations, we mined the Surveillance, Epidemiology, and End Results (SEER) database to construct the nomogram and used data from four academic hospitals for independent validation. This study contributes to providing more personalized guidance for patient care and improving patients' prognosis.

MATERIALS AND METHODS

Data Collection

In the present study, patients diagnosed with osteosarcoma between 2010 and 2016 were collected. The training group were extracted from the Surveillance, Epidemiology, and End Results (SEER) database with the SEER * Stat software version 8.3.6. And the third edition of the International Taxonomy of Oncology (ICDO-3), morphological code (9220) was used to identify osteosarcoma. The exclusion criteria of the training group were as follows: (1) patients with no positive pathology; (2) patients with unknown lymph node status and survival time; (3) more than one primary tumor.

Data of the validation group were obtained from four academic institutions, the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University, the Liuzhou People's Hospital affiliated to Guangxi Medical University, and the Xianyang Central Hospital. And during the period of investigation, each center was responsible for the acquisition of data by three investigators. Two investigators were responsible for data extraction and the accuracy check was conducted by the third investigator. The exclusion criteria were consistent with the training group. For multicenter data, the study was approved by the ethics review committee of four medical institutions in China, the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University, Liuzhou People's Hospital, and Xianyang Central Hospital (No. 2021-00-22) and was conducted in accordance with the guidelines of the Helsinki Declaration.

Demographic and clinical variables, including race, age, survival time, sex, primary site, grade, laterality, tumor size (T-stage), distance metastases (M-stage), surgery, radiation, chemotherapy, bone metastases and lung metastases were considered in this study.

Abbreviations: SEER, Surveillance, Epidemiology, and End Results database; ROC, receiver operating characteristic; DCA, decision curve analysis; CI, confidence interval; AUC, area under the curve; LNN, lymph node negative; LNM, lymph node metastases; ICD, International Classification of Diseases; SD, standard deviation; KM curves, Kaplan-Meier curves; OR, odds ratio.

Statistical Analysis

Continuous variables were expressed as mean ± standard deviation (SD), and categorical variables were expressed as frequency (proportions). The Student's *t*-test, Chi-square tests and Mann–Whitney tests were applied to continuous variables and categorical variables respectively with IBM SPSS Statistics version 26.0 (SPSS Inc., Chicago, Illinois, USA). Risk factors of the osteosarcoma were assessed using logistic regression. All analyzes were performed using R software version 3.6.2 (http://www.r-project.org) including multiple R packages (Including regplot, rms, rmda and pROC). P values < 0.05 were considered statistically significant, and confidence intervals (CIs) were expressed as 95% confidence levels.

Construction, Validation and Clinical Utility of a Nomogram

The following variables were included in the univariate logistic regression analysis: race, age, survival time, sex, primary site, grade, laterality, tumor size, lymphatic metastasis, surgery, radiation, chemotherapy, bone metastases and lung metastases. According to the result of the univariate logistic regression analysis with the P value < 0.05, we performed the multivariable logistic regression analysis. And the Nomogram was constructed based on the results of multivariable logistic regression analysis with the P value < 0.05. Calibration plot and receiver operating characteristic (ROC) curves were used to evaluate the prediction performance of the nomogram. The higher the area under the curves (AUC) of ROC indicated the better model performance. In addition, the decision curve analysis (DCA) was used to evaluate the clinical utility of nomogram in decision-making. Based on the established nomogram, an interactive convenient web calculator was provided (https://drliwenle.shinyapps.io/LMOOapp/).

RESULTS

Demographic Baseline Characteristics

As shown in **Table 1**, a total of 1,256 patients were enrolled. There was no statistically significant difference between the training group ($n = 1{,}144$) and validation group ($n = 112$) except the Race ($P < 0.001$) and Chemotherapy ($P = 0.017$).

Additionally, these patients were divided into two subgroups according to the LNM in **Table 2**. There were no significant differences in race, sex, laterality and radiation between the lymph node negative group (LNN, $n = 1{,}104$) and the lymph node positive (or unable to evaluate) group (LNP, $n = 152$). However, the other characteristics showed significant differences between the two groups.

TABLE 1 | Baseline data table of the training group and the validation group.

Variable	level	Overall (*N* = 1,256)	SEER data (Training group, *N* = 1,144)	Multicenter data (validation group, *N* = 112)	*p*
Race (%)	Black	168 (13.4)	168 (14.7)	0 (0.0)	<0.001
	Other	228 (18.2)	116 (10.1)	112 (100.0)	
	White	860 (68.5)	860 (75.2)	0 (0.0)	
Age [mean (SD)]	NA	33.31 (24.31)	33.47 (24.26)	31.62 (24.88)	0.443
Times [mean (SD)]	NA	29.93 (22.69)	29.91 (22.54)	30.10 (24.24)	0.933
Sex (%)	Female	573 (45.6)	521 (45.5)	52 (46.4)	0.936
	Male	683 (54.4)	623 (54.5)	60 (53.6)	
Primary.Site (%)	Axis bone	336 (26.8)	309 (27.0)	27 (24.1)	0.349
	Limb bone	817 (65.0)	738 (64.5)	79 (70.5)	
	Other	103 (8.2)	97 (8.5)	6 (5.4)	
Grade (%)	Moderately differentiated	41 (3.3)	41 (3.6)	0 (0.0)	0.124
	Poorly differentiated	302 (24.0)	279 (24.4)	23 (20.5)	
	Undifferentiated; anaplastic	560 (44.6)	511 (44.7)	49 (43.8)	
	Unknown	324 (25.8)	287 (25.1)	37 (33.0)	
	Well-differentiated	29 (2.3)	26 (2.3)	3 (2.7)	
Laterality (%)	Left	537 (42.8)	494 (43.2)	43 (38.4)	0.08
	Not a paired site	173 (13.8)	163 (14.2)	10 (8.9)	
	Right	546 (43.5)	487 (42.6)	59 (52.7)	
T (%)	T1	426 (33.9)	388 (33.9)	38 (33.9)	0.294
	T2	569 (45.3)	523 (45.7)	46 (41.1)	
	T3	42 (3.3)	35 (3.1)	7 (6.2)	
	TX	219 (17.4)	198 (17.3)	21 (18.8)	
M (%)	M0	976 (77.7)	892 (78.0)	84 (75.0)	0.547
	M1	280 (22.3)	252 (22.0)	28 (25.0)	
Surgery (%)	No	254 (20.2)	230 (20.1)	24 (21.4)	0.834
	Yes	1002 (79.8)	914 (79.9)	88 (78.6)	
Radiation (%)	No	1103 (87.8)	999 (87.3)	104 (92.9)	0.119
	Yes	153 (12.2)	145 (12.7)	8 (7.1)	
Chemotherapy (%)	No	274 (21.8)	260 (22.7)	14 (12.5)	0.017
	Yes	982 (78.2)	884 (77.3)	98 (87.5)	
Bone metastases (%)	No	1146 (91.2)	1044 (91.3)	102 (91.1)	0.981
	Unknown	52 (4.1)	47 (4.1)	5 (4.5)	
	Yes	58 (4.6)	53 (4.6)	5 (4.5)	
Lung metastases (%)	No	988 (78.7)	901 (78.8)	87 (77.7)	0.931
	Unknown	48 (3.8)	44 (3.8)	4 (3.6)	
	Yes	220 (17.5)	199 (17.4)	21 (18.8)	

Univariate and Multivariable Logistic Regression Results

According to the univariate logistics regression analysis, we identified 10 prognostic factors including age, survival time, primary site, laterality, tumor size, lymph metastasis, surgery, chemotherapy, bone metastases and lung metastases in the training set ($P < 0.05$) (**Table 3**). Then, based on the above result, applying the multivariable logistics regression analysis, we figured out five independent prognostic factors including T-stage [TX: odds ratio (OR) 3.602,95%CI 1.710–5.483, $P < 0.001$], M-stage (M1: OR = 2.890, 1.463–5.709, $P < 0.01$), surgery (Yes: OR = 0.418, 0.247–0.706, $P < 0.01$), Chemotherapy (Yes: OR = 0.475, 0.267–0.819, $P < 0.01$)and Lung metastases (Unknown: OR = 9.407, 1.955–45.261, $P < 0.01$) (**Table 3**).

Construction and Validation of Nomogram for Chondrosarcoma Patients

The nomogram contained five risk factors confirmed to be statistically significant by logistic regression analysis, including T-stage, M-stage, surgery, chemotherapy and lung metastases (**Figure 1A**). Calibration chart of nomogram showed a good consistency in the training and validation groups (**Figures 1B,C**). The AUC values of nomogram were 0.805 (95% CI 0.781–0.827) and 0.808 (95% CI 0.723–0.877) in the training group and

TABLE 2 | Patient baseline table of lymphatic metastases.

	Level	Overall (*N* = 1,256)	No (*N* = 1,104)	Yes/Unable to evaluate (*N* = 152)	*p*
Category (%)	Multicenter data (validation group)	112 (8.9)	93 (8.4)	19 (12.5)	0.133
	SEER data (Training group)	1144 (91.1)	1011 (91.6)	133 (87.5)	
Times [mean (SD)]	NA	29.93 (22.69)	30.95 (22.70)	22.45 (21.18)	<0.001
Race (%)	black	168 (13.4)	145 (13.1)	23 (15.1)	0.417
	Other	228 (18.2)	196 (17.8)	32 (21.1)	
	White	860 (68.5)	763 (69.1)	97 (63.8)	
Age [mean (SD)]	NA	33.31 (24.31)	32.48 (23.79)	39.35 (27.16)	0.001
Sex (%)	Female	573 (45.6)	498 (45.1)	75 (49.3)	0.37
	Male	683 (54.4)	606 (54.9)	77 (50.7)	
Primary.Site (%)	Axis bone	336 (26.8)	282 (25.5)	54 (35.5)	0.027
	Limb bone	817 (65.0)	732 (66.3)	85 (55.9)	
	Other	103 (8.2)	90 (8.2)	13 (8.6)	
Grade (%)	Moderately differentiated	41 (3.3)	39 (3.5)	2 (1.3)	<0.001
	Poorly differentiated	302 (24.0)	273 (24.7)	29 (19.1)	
	Undifferentiated; anaplastic	560 (44.6)	508 (46.0)	52 (34.2)	
	Unknown	324 (25.8)	257 (23.3)	67 (44.1)	
	Well-differentiated	29 (2.3)	27 (2.4)	2 (1.3)	
Laterality (%)	Left	537 (42.8)	483 (43.8)	54 (35.5)	0.104
	Not a paired site	173 (13.8)	146 (13.2)	27 (17.8)	
	Right	546 (43.5)	475 (43.0)	71 (46.7)	
T (%)	T1	426 (33.9)	400 (36.2)	26 (17.1)	<0.001
	T2	569 (45.3)	526 (47.6)	43 (28.3)	
	T3	42 (3.3)	36 (3.3)	6 (3.9)	
	TX	219 (17.4)	142 (12.9)	77 (50.7)	
M (%)	M0	976 (77.7)	883 (80.0)	93 (61.2)	<0.001
	M1	280 (22.3)	221 (20.0)	59 (38.8)	
Surgery (%)	No	254 (20.2)	179 (16.2)	75 (49.3)	<0.001
	Yes	1002 (79.8)	925 (83.8)	77 (50.7)	
Radiation (%)	No	1103 (87.8)	975 (88.3)	128 (84.2)	0.187
	Yes	153 (12.2)	129 (11.7)	24 (15.8)	
Chemotherapy (%)	No	274 (21.8)	208 (18.8)	66 (43.4)	<0.001
	Yes	982 (78.2)	896 (81.2)	86 (56.6)	
Bone metastases (%)	No	1146 (91.2)	1046 (94.7)	100 (65.8)	<0.001
	Unknown	52 (4.1)	14 (1.3)	38 (25.0)	
	Yes	58 (4.6)	44 (4.0)	14 (9.2)	

the validation group respectively (**Figures 2A,B**). Furthermore, the ROC curve demonstrated a superior performance of the nomogram compared to the single variable, including chemotherapy (AUC = 0.631, 95%CI 0.602 to 0.659), lung metastases (0.697, 0.669 to 0.723), M-stage (AUC = 0.592, 95%CI 0.563 to 0.621), surgery (AUC = 0.667, 95%CI 0.639 to 0.694) and T-stage (0.706, 95%CI 0.678 to 0.732). The statistical results of validation group were consistent with the training group as shown in **Table 4**. In addition, an online web calculator was designed (https://drliwenle.shinyapps.io/LMOOapp/).

Clinical Applicability of the Nomogram

The Kaplan-Meier survival curves of training group were plotted (**Figure 3A**). The results revealed that the overall survival (OS) significantly decreased in patients with LNP comparing with the LNN ($P < 0.001$). Moreover, the threshold about 0.1 to 0.9 had the maximum benefit range of the model as shown in the DCA curve (**Figures 3C,D**). The Kaplan-Meier survival curves of validation group displayed the same trend between the two groups ($P < 0.001$) (**Figure 3B**).

DISCUSSION

Osteosarcoma metastases, which are typically secondary to hematogenous dissemination and the occurrence of lymph system is extremely rare, have been identified to be significantly associated with poor prognosis (1–4, 7, 20, 21). Comparing to

TABLE 3 | Univariate and multifactorial logistic regression analysis of risk factors for Lymph node metastases in patients with osteosarcoma.

Variables	Univariate OR (95% CI)	p value	Multivariate OR (95% CI)	*p* value
Age (years)	1.011 (1.004–1.018)	<0.01	0.993 (0.982–1.004)	0.397
Survival time (month)	0.983 (0.974–0.992)	<0.001	0.998 (0.987–1.010)	0.788
Race				
White	Reference	Ref	Ref	Ref
Black	1.248 (0.766–3.033)	0.374	/	/
Other	0.993 (0.537–1.835)	0.982	/	/
Sex				
Male	Ref	Ref	Ref	Ref
Female	1.163 (0.810–1.671)	0.412	/	/
Primary site				
Limb bones	Ref	Ref	Ref	Ref
Axis of a bone	1.675 (1.133–2.478)	<0.05	0.946 (0.517–1.731)	0.857
Other	1.286 (0.671–2.466)	0.449	1.840 (0.806–4.198)	0.148
Grade				
Well-differentiated	Ref	Ref	Ref	Ref
Moderately differentiated	1.282 (0.110–14.893)	0.843	/	/
Poorly differentiated	2.569 (0.334–19.741)	0.364	/	/
Undifferentiated; anaplastic	2.473 (0.328–18.673)	0.380	/	/
Unknown	6.332 (0.840–47.705)	0.073	/	/
Laterality				
Left	Ref	Ref	Ref	Ref
Right	1.395 (0.930–2.091)	0.108	1.477 (0.920–2.371)	0.106
Other	1.848 (1.102–3.101)	<0.05	1.113 (0.536–2.310)	0.774
T				
T1	Ref	Ref	Ref	Ref
T2	1.121 (0.657–1.912)	0.675	0.960 (0.540–1.707)	0.889
T3	2.528 (0.900–7.101)	0.078	1.473 (0.476–4.559)	0.502
TX	7.933 (4.780–13.167)	<0.001	3.062 (1.710–5.483)	<0.001
M				
M0	Ref	Ref	Ref	Ref
M1	2.506 (1.710–3.673)	<0.001	2.890 (1.463–5.709)	<0.01
Surgery				
No	Ref	Ref	Ref	Ref
Yes	0.197 (0.135–0.287)	<0.001	0.418 (0.247–0.706)	<0.01
Radiation				
No	Ref	Ref	Ref	Ref
Yes	1.524 (0.936–2.481)	0.091	/	/
Chemotherapy				
No	Ref	Ref	Ref	Ref
Yes	0.289 (0.199–0.421)	<0.001	0.475 (0.276–0.819)	<0.01
Bone metastases				
No	Ref	Ref	Ref	Ref
Yes	3.260 (1.652–6.436)	<0.01	1.257 (0.570–2.772)	0.571
Unknown	32.490 (16.267–64.892)	<0.001	2.159 (0.483–9.659)	0.314
Lung metastases				
No	Ref	Ref	Ref	Ref
Yes	2.938 (1.887–4.573)	<0.001	0.902 (0.449–1.812)	0.771
Unknown	45.225 (21.360–95.754)	<0.001	9.407 (1.955–45.261)	<0.01

the most common lung metastases, osteosarcoma patients with lymph node involvement have a worse prognosis, suggesting that the invasion of the lymph nodes is an important indicator for the assessment of malignancy stage and the selection of a correct treatment protocol (22, 23).Meanwhile, studies points out that FDG-PET and ^{99m}Tc-labeled biomineralization

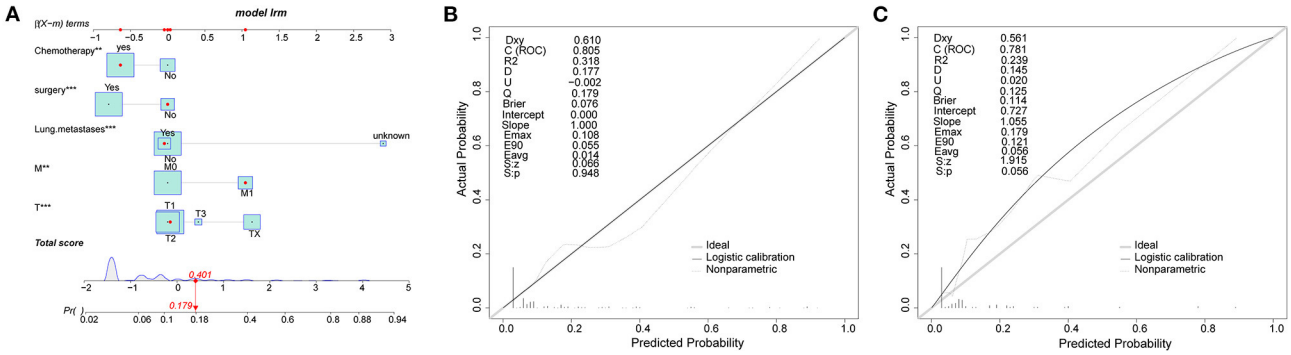


FIGURE 1 | (A) Nomogram for osteosarcoma patients. **(B,C)** are training cohorts and the validation cohorts calibration diagram respectively, which indicate good consistency.

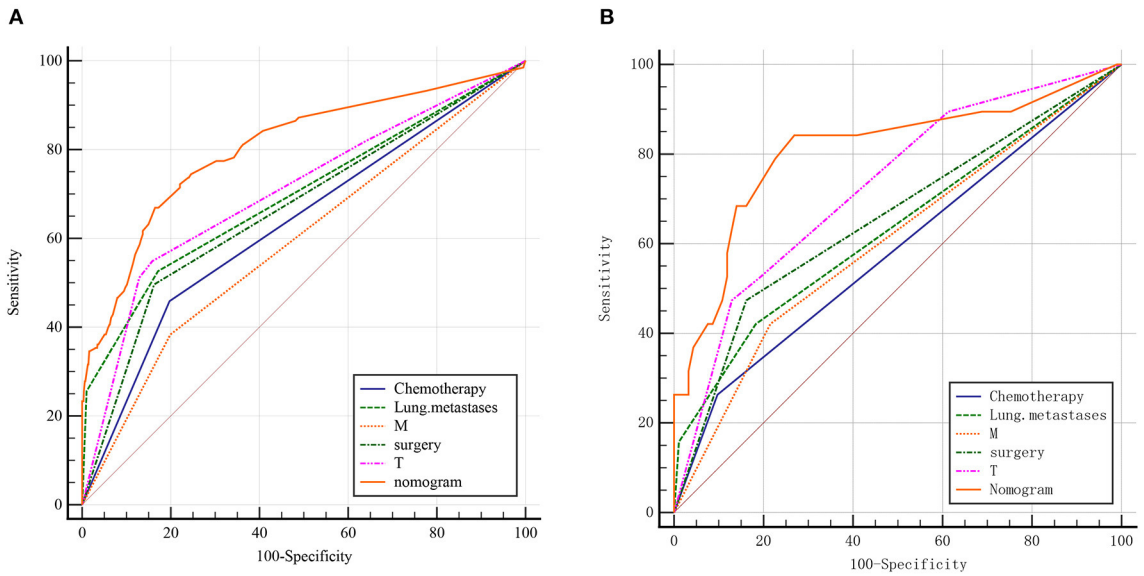


FIGURE 2 | ROC curves for the training and validation group. **(A)** training group; **(B)** validation group.

TABLE 4 | AUC of training group and validation group.

	SEER data (Training group)			Multicenter data (validation group)		
Variable	**AUC**	**SE**	**95% CI**	**AUC**	**SE**	**95% CI**
Chemotherapy	0.631	0.0226	0.602 to 0.659	0.583	0.0541	0.486 to 0.676
Lung metastases	0.697	0.0239	0.669 to 0.723	0.631	0.0644	0.535 to 0.721
M	0.592	0.0221	0.563 to 0.621	0.603	0.062	0.506 to 0.694
Surgery	0.667	0.0225	0.639 to 0.694	0.656	0.0619	0.561 to 0.743
T	0.706	0.0256	0.678 to 0.732	0.728	0.0614	0.635 to 0.807
Nomogram	0.805	0.0231	0.781 to 0.827	0.808	0.065	0.723 to 0.877

nanoprobe are effective in early diagnosis of metastatic lymph nodes in osteosarcoma (24–27). Therefore, real-time lymph node surveillance and radical treatment for osteosarcoma patients with a high risk of lymph node metastasis will improve patient survival (8). In this study, we identified five independent risk factors associated with LNM and provided a convenient nomogram prediction model and a web calculator on the basis of the model.

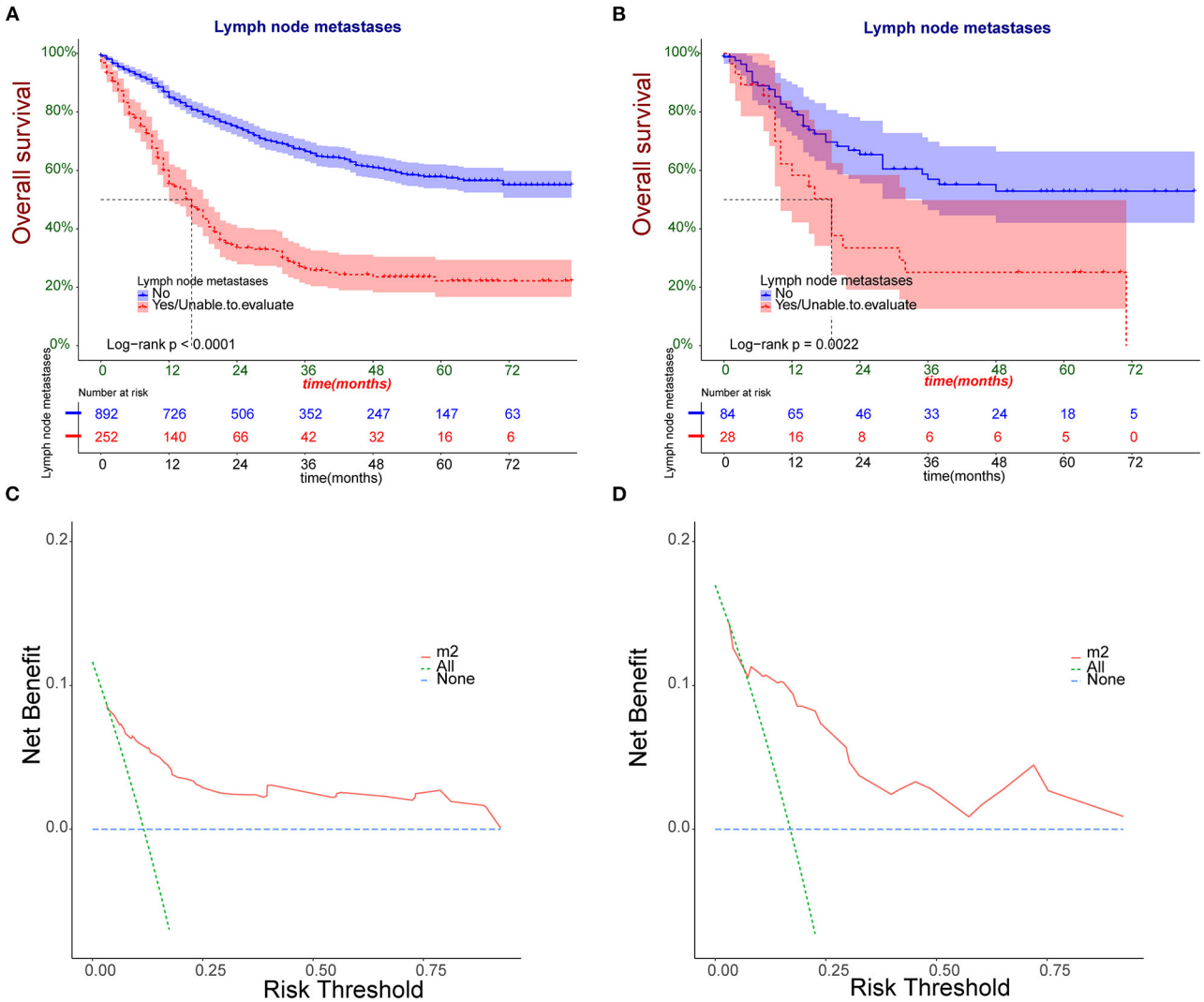


FIGURE 3 | (A,B) The Kaplan-Meier survival analysis patients with osteosarcoma in training and validation group. **(C,D)** Nomogram decision curve (DCA) of the training and validation group.

Surgery and chemotherapy are the most reliable and effective treatment options for prolonging the lives of patients and have been followed in several clinical trials (28–30). According to the multivariable logistic analysis, chemotherapy and surgery were crucial prognostic indicators. Surgery and chemotherapy contributed to improve patient prognosis, which were consistent with the result of the KM survival curve. This suggests that timely and effective treatment plays an important role in controlling lymph node metastasis and enhancing OS (15, 17, 31, 32). Unfortunately, the relationship between radiotherapy and lymph node metastasis has not been proved. A possible reason for these results would be the uncertainty of radiation therapy in cancer treatment along with the development of radiation-associated osteosarcoma make the safety of chemotherapy need to be further ensured (33, 34).

Due to the lack of lymphatic drainage in normal cortical bone and spongy bone, LNM is rare in bone sarcomas (35, 36).Regional lymph node involvement in osteosarcoma may be owing to the infiltration of the enlarged tumor parenchyma into the periphery, such as the joint capsule or synovium, leading to dissemination into the lymphatic system (37, 38).Our study reported T-stage as a significant predictor of LNM, which was also consistent with previous findings (7).Meanwhile, because of the fact that the peak incidence of the osteosarcoma is 15–19 years of age, more than 80% of patients achieve limb salvage through surgery (4, 28, 38).In this respect, the surgeons need pay more attention to the status of the regional lymph nodes during the resection of larger osteosarcoma in order to eradicate the sarcoma and preserve the limb. M-stage indicates distant metastasis in osteosarcoma, and it is also a risk factor associated with LNM. In the study by Thampi et al. osteosarcoma lymphatic

metastasis was significantly associated with distant metastasis (7).Furthermore, since the lung was the most important target organ for distant metastasis in osteosarcoma, we separated this variable from distant metastasis to emphasize the significant role played by lung metastasis in assessing lymph node status. Patients with lung metastases express the characteristic biomarker, such as KEAP, Matrix-Gla and Rab22a (39–41). Considering the intense correlation we found between lung metastasis and lymph node involvement, these reported biomarkers would be triggers for LNM. Therefore, this provides inspiration to further molecular biology studies focusing on lymph node metastasis in osteosarcoma.

We constructed a novel nomogram to assess the risk of developing lymph node metastasis in osteosarcoma, and the discriminatory of any individual predictor was inferior to that of nomogram, suggesting that the nomogram model indicated promising prospects for tumor surveillance and clinical decision making. Although some predictive nomograms have been reported in previous studies, our study complements previous work. Compared to Dong et al.'s study, external independent validation consisting of multiple academic centers is a prominent feature of this study, and the inclusion of multiple ethnic groups enhances the credibility of the results (8). Moreover, we provide a convenient and digital prediction tool for users. By analyzing the clinical characteristics and associated risk factors, we improve the prediction of lymph node metastasis risk in osteosarcoma and provide a basis for individualized treatment and follow-up strategies. The web-based calculator constructed in this study is an easy-to-use clinical tool that helps to promote personalized treatment.

Finally, one obvious limitation in this study was that the statistically significant difference between the training group and validation groups in chemotherapy may have an influence on the results. Another limitation was the lack of a more thoroughly analysis due to the inadequacy of systematic and prospective data.

CONCLUSION

In conclusion, we constructed a novel nomogram model to predict risk factors for osteosarcoma patients developing LNM, including T-stage, M-stage, surgery, chemotherapy and lung metastases based on epidemiological characteristics obtained from the SEER database and the multicenter hospitals. By combining DCA curve, ROC curve, KM curve, web calculator and external validation, our nomogram provided an accurate assessment for individualized risk of lymph node metastasis which was helpful for clinicians to make better surgery decisions.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

ETHICS STATEMENT

The SEER database is a comprehensive data source developed based on population data and updated annually since its launch in 1973. It is public and identifiably accessible that data analysis is treated as non-human subjects by the Office for Human Research Protections. As such, no institutional review board approval and informed consent were required. For multicenter data, the study was approved by the ethics review committee of four medical institutions in China, the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University, 291 Liuzhou People's Hospital, and Xianyang Central Hospital (No. 2021-00-22) and was conducted in accordance with the guidelines of the Helsinki.

AUTHOR CONTRIBUTIONS

CY, RW, QL, and XL designed the study. WeL and SD performed the study, analyzed the data, and wrote the manuscript. BW and HW provided the expert consultations and clinical suggestions. CX and KZ conceived of the study, participated in its design and coordination. WaL and ZH helped to draft the manuscript. All authors reviewed the final version of the manuscript.

ACKNOWLEDGMENTS

We thank China's multicenter for providing us individual patient data. The multicenter data were obtained from four medical institutions in China, the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University, Liuzhou People's Hospital, and Xianyang Central Hospital.

REFERENCES

1. Marina N, Smeland S, Bielack S, Bernstein M, Jovic G, Hook J, et al. "MAPIE vs MAP as postoperative chemotherapy in patients with a poor response to preoperative chemotherapy for newly-diagnosed osteosarcoma: results from EURAMOS-1 (Paper 032)," in *Presented at Connective Tissue Oncology Society* (2014).
2. Tian Z, Niu X, Yao W. Receptor tyrosine kinases in osteosarcoma treatment: which is the key target? *Front Oncol.* (2020) 10:1642. doi: 10.3389/fonc.2020.01642
3. Kager L, Zoubek A, Pötschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols. *J Clin Oncol.* (2003) 21:2011–8. doi: 10.1200/JCO.2003.08.132
4. Hattori H, Yamamoto K. Lymph node metastasis of osteosarcoma. *J Clin Oncol.* (2012) 30:345–9. doi: 10.1200/JCO.2012.42.3384
5. Kempf-Bielack B. Osteosarcoma relapse after combined modality therapy: an analysis of unselected patients in the Cooperative Osteosarcoma Study Group (COSS). *J Clin Oncol.* (2005) 23:559–68. doi: 10.1200/JCO.2005.04.063
6. Zheng B, Zhou C, Qu G, Ren C, Yue B. VEGFR2 promotes metastasis and PD-L2 expression of human osteosarcoma cells by activating the STAT3 and RhoA-ROCK-LIMK2 pathways. *Front Oncol.* (2020) 10:543562. doi: 10.3389/fonc.2020.543562

7. Thampi S, Matthay KK, Goldsby R, Dubois SG. Adverse impact of regional lymph node involvement in osteosarcoma. *Eur J Cancer.* (2013) 49:3471–6. doi: 10.1016/j.ejca.2013.06.023
8. Dong Y, Wu W, Kang H, Xiong W, Li F. Risk factors of regional lymph node (RLN) metastasis among patients with bone sarcoma and survival of patients with RLN-positive bone sarcoma. *Ann Transl Med.* (2021) 9:48. doi: 10.21037/atm-20-4681
9. Kumar R, Tripathi R, Marchang N, Srivastava G, Gadekallu TR, Xiong NN. A secured distributed detection system based on IPFS and blockchain for industrial image and video data security. *J Parallel Distributed Comput.* (2021) 152:128–43. doi: 10.1016/j.jpdc.2021.02.022
10. Agrawal S, Sarkar S, Srivastava G, Maddikunta P, Gadekallu TR. Genetically optimized prediction of remaining useful life. *Sustain Comput.* (2021) 31:100565. doi: 10.1016/j.suscom.2021.100565
11. Naeem A, Javed AR, Rizwan M, Abbas S, Gadekallu TR. DARE-SEP: a hybrid approach of distance aware residual energy-efficient SEP for WSN. *IEEE Trans Green Commun Netw.* (2021) 5:611–21. doi: 10.1109/TGCN.2021.3067885
12. Gadekallu TR, Alazab M, Kaluri R, Maddikunta P, Parimala M. Hand gesture classification using a novel CNN-crow search algorithm. *Complex Intell Syst.* (2021) 7:1855–68. doi: 10.1007/s40747-021-00324-x
13. Gadekallu TR, Manoj MK, Sivarama KS, Kumar N, Bhattacharya S. Blockchain based attack detection on machine learning algorithms for IoT based E-health applications. *IEEE Internet Things Mag.* (2020) 4:30–33. doi: 10.1109/IOTM.1021.2000160
14. Tang Z, Zhu R, Lin P, He J, Wang H, Huang Q, et al. A hardware friendly unsupervised memristive neural network with weight sharing mechanism. *Neurocomputing.* (2019) 332:193–202. doi: 10.1016/j.neucom.2018.12.049
15. Wu EQ, Hu D, Deng PY, Tang Z, Ren H. Nonparametric bayesian prior inducing deep network for automatic detection of cognitive status. *IEEE Trans Cybern.* (2020) 51:5483–96. doi: 10.1109/TCYB.2020.2977267
16. Tang Z, Chen Y, Ye S, Hu R, Chang S. Fully Memristive spiking-neuron learning framework and its applications on pattern recognition and edge detection. *Neurocomputing.* (2020) 403:80–7. doi: 10.1016/j.neucom.2020.04.012
17. Wu EQ, Deng PY, Qu XY, Tang Z, Sheng RSF. Detecting fatigue status of pilots based on deep learning network using EEG signals. *IEEE Trans Cogn Dev Syst.* (2020) 13:575–85. doi: 10.1109/TCDS.2019.2963476
18. Tang Z, Zhu R, Hu R, Chen Y, Chang S. A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors. *IEEE Trans Cogn Dev Syst.* (2020) 13:645–56. doi: 10.1109/TCDS.2020.3003377
19. Balachandran VP, Gonen M, Smith JJ, Dematteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* (2015) 16:e173–e80. doi: 10.1016/S1470-2045(14)71116-7
20. Meazza C, Scanagatta P. Metastatic osteosarcoma: a challenging multidisciplinary treatment. *Expert Rev Anticancer Ther.* (2016) 16:543–56. doi: 10.1586/14737140.2016.1168697
21. Wang T, Quan Y, Shen XS, Gadekallu TR, Wang W, Dev K, et al. Privacy-enhanced retrieval technology for the cloud-assisted internet of things. *IEEE Trans Ind Inform.* (2021) 14:1551–3203. doi: 10.1109/TII.2021.3103547
22. Alitalo K. The lymphatic vasculature in disease. *Nat Med.* (2011) 17:1371–80. doi: 10.1038/nm.2545
23. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet Things J.* (2021) 1–1. doi: 10.1109/JIOT.2021.3117762
24. Xu Z, Wang Y, Han J, Xu Q, Ren J, Xu J, et al. Noninvasive multimodal imaging of osteosarcoma and lymph nodes using a 99mTc-labeled biomineralization nanoprobe. *Anal Chem.* (2018) 90:4529–34. doi: 10.1021/acs.analchem.7b04925
25. Volker T, Denecke T, Steffen I, Misch D, Schonberger S, Plotkin M, et al. Positron emission tomography for staging of pediatric sarcoma patients: results of a prospective multicenter trial. *J Clin Oncol.* (2007) 25:5435. doi: 10.1200/JCO.2007.12.2473
26. Wang W, Fida MH, Lian Z, Yin Z, Pham QV, Gadekallu TR, et al. Secure-enhanced federated learning for AI-empowered electric vehicle energy prediction. *IEEE Consum Electron Mag.* (2021). doi: 10.1109/MCE.2021.3116917
27. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inform.* (2021). doi: 10.1109/JBHI.2021.3112693
28. Ritter J, Bielack SS. Osteosarcoma. *Ann Oncol.* (2010) 21(Suppl 7):vii320–5. doi: 10.1093/annonc/mdq276
29. Pramanik R, Agarwala S, Gupta YK, Thulkar S, Bakhshi S. Metronomic chemotherapy vs best supportive care in progressive pediatric solid malignant tumors: a randomized clinical trial. *Jama Oncol.* (2017) 3:1222–7. doi: 10.1001/jamaoncol.2017.0324
30. Piperno-Neumann S, Le Deley MC, Rédini F, Pacquement H, Marec-Bérard P, Petit P, et al. Zoledronate in combination with chemotherapy and surgery to treat osteosarcoma (OS2006): a randomised, multicentre, open-label, phase 3 trial. *Lancet Oncol.* (2016) 17:1070–80. doi: 10.1016/S1470-2045(16)30096-1
31. Huang Z, Hu C, Liu K, Yuan L, Hu C. Risk factors, prognostic factors, and nomograms for bone metastasis in patients with newly diagnosed infiltrating duct carcinoma of the breast: a population-based study. *BMC Cancer.* (2020) 20:1145. doi: 10.1186/s12885-020-07635-1
32. Wu EQ, Zhou MC, Hu D, Zhu L, Ren H. Self-paced dynamic infinite mixture model for fatigue evaluation of pilots' brains. *IEEE Trans Cybern.* (2020) 2168–267. doi: 10.1109/TCYB.2020.3033005
33. Siontis BL, Mchugh JB, Roberts E, Zhao L, Chugh R. Differential outcomes and biologic markers of radiation-associated vs. sporadic osteosarcoma: a single-institution experience. *Front Oncol.* (2020) 9:1523. doi: 10.3389/fonc.2019.01523
34. Wu EQ, Zhu LM, Li GJ, Li HJ, Zhou GR. Nonparametric hierarchical hidden semi-markov model for brain fatigue behavior detection of pilots during flight. *IEEE Trans Intell Transport Syst.* (2021) 1–12. doi: 10.1109/TITS.2021.3052801
35. Cleary MX, Fayad LM, Ahlawat S. Popliteal lymph nodes in patients with osteosarcoma: are they metastatic? *Skeletal Radiol.* (2020) 49:1807–17. doi: 10.1007/s00256-020-03498-6
36. Wu EQ, Lin CT, Zhu LM, Tang ZR, Jie YW, Zhou GR. Fatigue detection of pilots' brain through brains cognitive map and multilayer latent incremental learning model. *IEEE Trans Cybern.* (2021) 1–13. doi: 10.1109/TCYB.2021.3068300
37. Dirik Y, Çinar A, Yumrukçal F, Eralp L. Popliteal lymph node metastasis of tibial osteoblastic osteosarcoma. *Int J Surg Case Rep.* (2014) 5:840–4. doi: 10.1016/j.ijscr.2014.09.029
38. Bielack SS, Kempf-Bielack B, Delling G, Exner GU, Flege S, Helmke K, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol.* (2002) 20:776–90. doi: 10.1200/JCO.2002.20.3.776
39. Xu H, Zhu X, Bao H, Shek TW, Huang Z, Wang Y, et al. Genetic and clonal dissection of osteosarcoma progression and lung metastasis. *Int J Cancer.* (2018) 143:1134—42. doi: 10.1002/ijc.31389
40. Zandueta C, Ormazábal C, Perurena N, Martínez-Canarias S, Zalacaín M, Julián MS, et al. Matrix-Gla protein promotes osteosarcoma lung metastasis and associates with poor prognosis. *J Pathol.* (2016) 239:438–49. doi: 10.1002/path.4740

41. Liao D, Zhong L, Yin J, Zeng C, Wang X, Huang X, et al. Chromosomal translocation-derived aberrant Rab22a drives metastasis of osteosarcoma. *Nat Cell Biol*. (2020) 22:868–81doi: 10.1038/s41556-020-0522-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Dong, Wang, Wang, Xu, Zhang, Li, Hu, Li, Liu, Wu and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fpubh.2021.812395

A Web-Based Prediction Model for Overall Survival of Elderly Patients With Malignant Bone Tumors: A Population-Based Study

Jie Tang[1†], *JinKui Wang*[2†] *and Xiudan Pan*[1*]

[1] *Department of Biostatistics and Epidemiology, School of Public Health, Shenyang Medical College, Shenyang, China,*
[2] *Department of Orthopedics, Children's Hospital of Chongqing Medical University, Chongqing, China*

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Sharnil Pandya, Symbiosis International University, India
Abdul Rehman Javed, Air University, Pakistan
Milad Janalipour, K.N.Toosi University of Technology, Iran

***Correspondence:**
Xiudan Pan
xiudanp@aliyun.com

[†]*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *10 November 2021*
Accepted: *13 December 2021*
Published: *11 January 2022*

Citation:
Tang J, Wang J and Pan X (2022) A Web-Based Prediction Model for Overall Survival of Elderly Patients With Malignant Bone Tumors: A Population-Based Study. Front. Public Health 9:812395. doi: 10.3389/fpubh.2021.812395

Background: Malignant bone tumors (MBT) are one of the causes of death in elderly patients. The purpose of our study is to establish a nomogram to predict the overall survival (OS) of elderly patients with MBT.

Methods: The clinicopathological data of all elderly patients with MBT from 2004 to 2018 were downloaded from the SEER database. They were randomly assigned to the training set (70%) and validation set (30%). Univariate and multivariate Cox regression analysis was used to identify independent risk factors for elderly patients with MBT. A nomogram was built based on these risk factors to predict the 1-, 3-, and 5-year OS of elderly patients with MBT. Then, used the consistency index (C-index), calibration curve, and the area under the receiver operating curve (AUC) to evaluate the accuracy and discrimination of the prediction model was. Decision curve analysis (DCA) was used to assess the clinical potential application value of the nomogram. Based on the scores on the nomogram, patients were divided into high- and low-risk groups. The Kaplan-Meier (K-M) curve was used to test the difference in survival between the two patients.

Results: A total of 1,641 patients were included, and they were randomly assigned to the training set (N = 1,156) and the validation set (N = 485). The univariate and multivariate analysis of the training set suggested that age, sex, race, primary site, histologic type, grade, stage, M stage, surgery, and tumor size were independent risk factors for elderly patients with MBT. The C-index of the training set and the validation set were 0.779 [0.759–0.799] and 0.801 [0.772–0.830], respectively. The AUC of the training and validation sets also showed similar results. The calibration curves of the training and validation sets indicated that the observed and predicted values were highly consistent. DCA suggested that the nomogram had potential clinical value compared with traditional TNM staging.

Conclusion: We had established a new nomogram to predict the 1-, 3-, 5-year OS of elderly patients with MBT. This predictive model can help doctors and patients develop treatment plans and follow-up strategies.

Keywords: nomogram, elderly patients, malignant bone tumors, overall survival, SEER

INTRODUCTION

The most common malignant bone tumors (MBT) in elderly patients include osteosarcoma, chondrosarcoma, and chord sarcoma. Osteosarcoma is the most common malignant bone tumor in children and adolescents. About 60% of osteosarcoma patients occur under 20 years of age, and ∼10% of patients occur over 60 years of age (1). Osteosarcoma usually occurs in the metaphysis of long tubular bones, most commonly around the knee joint. Osteosarcoma in adults can affect the axial and maxillofacial bones (2). Patients with osteosarcoma over 65 years of age are often secondary to Paget's disease (3). At the time of diagnosis of osteosarcoma, 10–20% of patients have already developed distant metastases, most of which are lung metastases. The prognosis of patients with osteosarcoma metastasis is inferior, and the overall 5-year survival rate is 20–30% (4). Chondrosarcoma accounts for about 9.2% of all primary MBT, with an annual incidence of 1/200,000 and an average age of onset of about 50 years old (5, 6). It is the second most common bone malignant tumor after osteosarcoma. Recent studies on chondrosarcoma report that the 5-year survival rate is about 75.2%, and the 10-year survival rate is about 70% (7, 8). Chondrosarcoma is mainly of the conventional type, and only 8–10% are of the unconventional kind. There is a big difference in survival and prognosis (9, 10). Chord sarcoma accounts for 1% to 4% of primary bone tumors, and the ratio of males to females is about 1.8:1. It can occur at any age, including children and adolescents, and is most common in patients between 50 and 60 years old (30%). Chord sarcoma grows slowly and occurs more frequently in the sacrococcygeal region (50–60%), skull base (25–35%), cervical spine (10%), thoracolumbar spine (5%), and rarely occurs in areas other than the axial bone (11). Although chord sarcoma is low to moderately malignant, 8% to 43% of patients still have distant metastasis (pulmonary metastasis is the most common), and the prognosis is poor. The overall 5-year survival rate of patients with metastatic chord sarcoma is only about 50% (11, 12).

Studies have reported that the main prognostic factors of osteosarcoma include tumor site, size, age, metastasis, metastasis site, chemotherapy, and surgery (13, 14). The significant prognostic factors of chondrosarcoma include primary tumor, histological type, tumor site, histological grade, and tumor size (15–17). One study showed that age, distant metastasis, and surgery are risk factors for chord sarcoma. However, there is no study on the prognostic factors of MBT in the elderly. There are very few elderly patients with MBT, and single-center studies cannot provide accurate estimates of prognostic factors.

Artificial intelligence has been widely used in human health care. Hassan et al. (18) summarized machine learning to predict the occurrence of sepsis to prevent patients from developing severe sepsis. Alhazmi et al. (19) used machine learning to predict the risk of oral cancer. Cho et al. (20) conducted a systematic review of the literature on machine learning to predict the risk of cancer brain metastasis.

The nomogram is a numerical model that predicts an outcome event by using estimated values generated based on various variables (21, 22). The construction of the nomogram is based on big clinical data, and a small number of cases may cause inaccurate predictions. The United States Surveillance, Epidemiology, and End Results (SEER) project is a cancer data center established in the United States in 1973, collecting information on cancer patients in 18 registries (23). The SEER database can provide enough cases to build a nomogram prediction model. Accurate prediction of survival time can help doctors better monitor patients. The prognostic factors of cancer-specific survival in elderly patients with bone tumors are not yet fully understood. Our purpose is to explore the prognostic factors affecting the survival of elderly patients with malignant bone tumors. Accurate prediction of survival time can help doctors better monitor patients. The prognostic factors of cancer-specific survival in elderly patients with bone tumors are not yet fully understood. Our purpose is to explore the prognostic factors affecting the survival of elderly patients with malignant bone tumors. We aim to construct a nomogram to predict the survival and prognosis of elderly patients with MBT to help doctors and patients formulate clinical treatment plans and follow-up strategies. After the model was established, we also conducted a series of validations on the prediction model's performance.

METHODS

Data Source and Data Extraction

The clinicopathological information of all patients over 60 years old with chondrosarcoma, osteosarcoma, and chord sarcoma from 2004 to 2018 was extracted from the SEER database. The extracted information includes age, gender, race, primary sit, histological grade, histological type, tumor size, TNM stage, surgery, radiotherapy, chemotherapy, and follow-up information. The selection criteria are: (1) age $\geq$ 60 years old; (2) histological classification includes: chondrosarcoma, osteosarcoma, and chord sarcoma. The exclusion criteria are: (1) survival time is less than one month; (2) tumor size is unknown; (3) TNM stage is unknown (4) surgical method is unknown; (5) tumor stage is unknown. The flow chart of patient screening is shown in **Figure 1**.

The age of all patients was divided into five groups, including 60–64, 65–69, 70–74, 75–79, $\geq$80 years old. The race was divided into three categories: white, black, and others (American Indian/AK Indian, Asian/Pacific Islander). The tumor grade was divided into four grades: well-differentiated, moderately differentiated, poorly differentiated, and undifferentiated, and some were unknown graded. The tumor stage was divided into three categories, including localized, regional, and distant. Surgical methods included no surgery, partial resection, radical resection, and amputation.

Nomogram Construction and Validation

All patients were randomly divided into a training set (70%) and a validation set (30%). Univariate and multivariate Cox regression models were used to determine independent risk factors for patients. These risk factors were included in the nomogram to predict the 1-, 3-, and 5-year OS of elderly patients with MBT. The consistency index (C-index) was used to test the discrimination of the nomogram. The area under the receiver

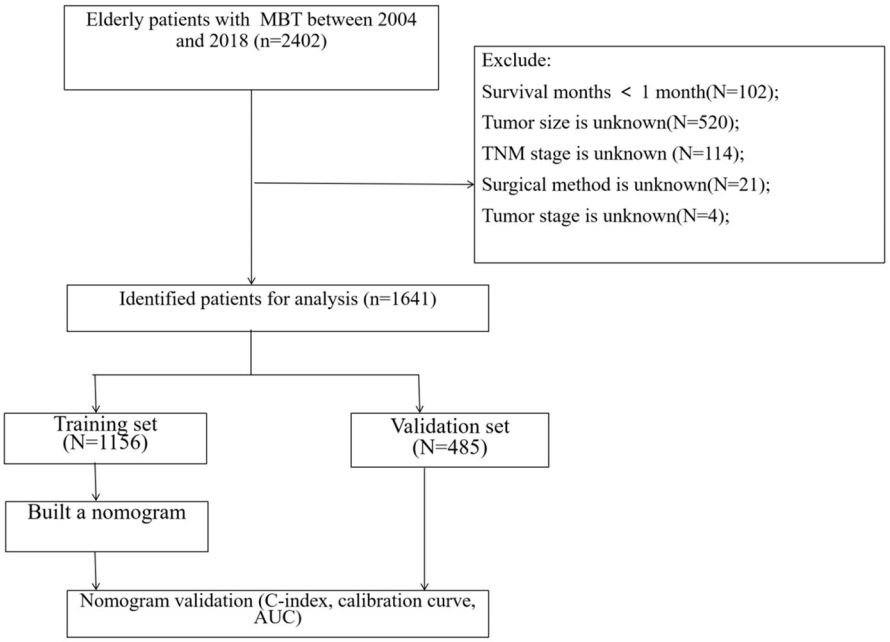


FIGURE 1 | The flowchart of including and dividing patients.

operating curve (AUC) was also used to evaluate the training and validation set's accuracy and discrimination. A calibration curve of 1,000 bootstrap weight samples was used to test the accuracy of the prediction model.

Clinical Utility

Decision curve analysis (DCA) of the training set and validation set was used to evaluate the clinical value of the nomogram. A new algorithm assesses the model's weight by calculating the net benefit under the risk threshold (24). We also compared the DCA benefit of the nomogram and traditional TNM staging. According to the nomogram score, all patients were divided into low-risk and high-risk groups. Kaplan-Meier curve and the log-rank test were used to test the difference in survival between the two groups. We also compared the survival differences of patients in different risk groups under different surgical methods.

Statistical Analysis

All statistical analyses were performed using SPSS version 26.0 and R software version 4.1.0. The median (inter-quartile range) was used to describe the data that did not follow the normal distribution for measurement data. The non-parametric test (U-test) was used to analyze the difference. For count data, used frequency (%) to describe, used chi-square analysis and non-parametric *U*-test for analysis. Univariate and multivariate Cox regression models were used to analyze survival risk factors. This study believes that $P<0.05$ is statistically significant, and this test is two-sided.

RESULTS

Clinical Features

A total of 1,641 patients were included, and they were randomly assigned to the training set ($N = 1{,}156$) and the validation set ($N = 485$). The clinical-pathological data of the patient is shown in **Table 1**. There were 887 males (54.1%) and 1,424 whites (86.8%). There were 958 cases (58.4%) of chondrosarcoma, 328 cases (20.0%) of osteosarcoma, and 355 cases (21.6%) of chord sarcoma. Patients with tumor grade I, II, III, and IV were 229 (14.0%), 404 (24.6%), 245 (14.9%), and 259 (15.8%), respectively; 504 (30.7%) patients were unknown grades. In the tumor stage, 713 cases (43.4%) were localized, 696 cases (42.4%) were regional, and 232 cases (14.1%) were distant. There were 336 (20.5%) without surgery, 458 (27.9%) with partial resection, 661 (40.3%) with radical resection, and 186 (11.3%) with amputation. Tumor primary sites included 623 cases (38.0%) in limbs, 231 cases (14.1%) in skull, 134 (8.17%) in spine, 246 cases (15.0%) in thorax, and 407 cases (24.8%) in pelvis. There is no significant difference between the clinical-pathological data of the patients in the training set and the validation set.

Univariate and Multivariate Cox Regression Analysis

We used univariate Cox regression analysis to confirm the risk factors for the prognosis initially. The selected risk factors were then included in the multivariate Cox regression analysis. In the end, we found that age, sex, race, primary site, histologic type, grade, stage, M stage, surgery, and tumor size were independent risk factors for elderly patients with MBT. The independent

TABLE 1 | Clinicopathological characteristics of elderly patients with MBT.

	Total	Training set	Validation set	*P*
	N = 1,641	*N* = 1,156	*N* = 485	
Age				0.054
60–64	416 (25.4%)	276 (23.9%)	140 (28.9%)	
65–69	376 (22.9%)	285 (24.7%)	91 (18.8%)	
70–74	338 (20.6%)	233 (20.2%)	105 (21.6%)	
75–79	223 (13.6%)	155 (13.4%)	68 (14.0%)	
≥80	288 (17.6%)	207 (17.9%)	81 (16.7%)	
Race				0.141
White	1,424 (86.8%)	1,015 (87.8%)	409 (84.3%)	
Black	94 (5.73%)	63 (5.45%)	31 (6.39%)	
Other	123 (7.50%)	78 (6.75%)	45 (9.28%)	
Sex				0.716
Male	887 (54.1%)	621 (53.7%)	266 (54.8%)	
Female	754 (45.9%)	535 (46.3%)	219 (45.2%)	
Year of diagnosis				0.710
2004–2010	613 (37.4%)	428 (37.0%)	185 (38.1%)	
2010–2018	1,028 (62.6%)	728 (63.0%)	300 (61.9%)	
Histologic type				0.258
Chondrogenic sarcoma	958 (58.4%)	687 (59.4%)	271 (55.9%)	
Osteogenic sarcoma	328 (20.0%)	231 (20.0%)	97 (20.0%)	
Chordal sarcoma	355 (21.6%)	238 (20.6%)	117 (24.1%)	
Grade				0.278
I	229 (14.0%)	153 (13.2%)	76 (15.7%)	
II	404 (24.6%)	295 (25.5%)	109 (22.5%)	
III	245 (14.9%)	181 (15.7%)	64 (13.2%)	
IV	259 (15.8%)	182 (15.7%)	77 (15.9%)	
Unknown	504 (30.7%)	345 (29.8%)	159 (32.8%)	
Stage				0.174
Localized	713 (43.4%)	503 (43.5%)	210 (43.3%)	
Regional	696 (42.4%)	501 (43.3%)	195 (40.2%)	
Distant	232 (14.1%)	152 (13.1%)	80 (16.5%)	
Primary site				0.123
Limb	623 (38.0%)	444 (38.4%)	179 (36.9%)	
Cranial	231 (14.1%)	161 (13.9%)	70 (14.4%)	
Spine	134 (8.17%)	103 (8.91%)	31 (6.39%)	
Thoracic	246 (15.0%)	179 (15.5%)	67 (13.8%)	
Pelvic	407 (24.8%)	269 (23.3%)	138 (28.5%)	
T				0.141
T1	940 (57.3%)	680 (58.8%)	260 (53.6%)	
T2	652 (39.7%)	446 (38.6%)	206 (42.5%)	
T3	38 (2.32%)	24 (2.08%)	14 (2.89%)	
T4	11 (0.67%)	6 (0.52%)	5 (1.03%)	
N				1.000
N0	1,605 (97.8%)	1,131 (97.8%)	474 (97.7%)	
N1	36 (2.19%)	25 (2.16%)	11 (2.27%)	
M				0.899
M0	1,486 (90.6%)	1,048 (90.7%)	438 (90.3%)	
M1	155 (9.45%)	108 (9.34%)	47 (9.69%)	
Tumor size(median [IQR])	70.0 [44.0, 112.0]	70.0 [43.0, 112.0]	71.0 [45.0, 111.0]	0.3598
Chemotherapy				0.975
No/Unknown	1,395 (85.0%)	982 (84.9%)	413 (85.2%)	
Yes	246 (15.0%)	174 (15.1%)	72 (14.8%)	

(Continued)

TABLE 1 | Continued

	Total	Training set	Validation set	*P*
	N = 1,641	*N* = 1,156	*N* = 485	
Radiation				0.986
No/Unknown	1,227 (74.8%)	865 (74.8%)	362 (74.6%)	
Yes	414 (25.2%)	291 (25.2%)	123 (25.4%)	
Surgery				0.767
No	336 (20.5%)	236 (20.4%)	100 (20.6%)	
Partial resection	458 (27.9%)	325 (28.1%)	133 (27.4%)	
Radical resection	661 (40.3%)	470 (40.7%)	191 (39.4%)	
Amputation	186 (11.3%)	125 (10.8%)	61 (12.6%)	

risk factors screened by univariate and multivariate analysis are shown in **Table 2**.

Nomogram Construction and Validation

A nomogram was constructed based on the identified independent risk factors to predict the 1-, 3-, and 5-year OS of elderly patients with MBT (**Figure 2**). It can be seen from the nomogram that tumor size has the most significant impact on OS, followed by primary tumor site, grade, and surgery. The calibration curves of the training set and the validation set showed that the observed value and the predicted value were highly consistent, indicating the accuracy of the nomogram prediction (**Figure 3**). The C-index of the training set and the validation set were 0.779 [0.759–0.799] and 0.801 [0.772–0.830], respectively, indicating that the nomogram had good discrimination. The AUC of the training set and the validation set also showed similar results (**Figure 4**).

Clinical Application of the Nomogram

DCA suggested that nomogram had potential clinical value compared with traditional TNM staging (**Figure 5**). Based on the score of each patient in the nomogram, the patients were divided into a low-risk group (total score $\leq$ 149.4) and a high-risk group (total score $>$ 149.4). Apparent differences in survival were observed from the K-M curves of the training set and validation sets' K-M curves (**Figure 6**). The K-M curve indicated that the nomogram has an excellent discriminating ability. We found that patients in the high-risk group who chose partial resection had a higher survival rate according to risk groups. At the same time, there is no significant difference in the prognosis of patients in the low-risk group with various treatment methods (**Figure 7**).

Online Application for OS Prediction

We have developed a web calculator based on the nomogram, which can be accessed at https://jietang.shinyapps.io/DynNomapp/. Enter the clinicopathological characteristics of the patient on this web page to get the corresponding survival probability. This calculation is easy to use and friendly to both patients and doctors.

DISCUSSION

Cancer is one of the most common causes of death in elderly patients. The morbidity and mortality of bone tumors are high. A study has reported 3,600 new bone tumor patients in the United States each year and 1,720 deaths from bone tumors (25). Although MBT is not common, their mortality is still high. The three most common MBT in elderly patients include chondrosarcoma, osteosarcoma, and chord sarcoma, of which chondrosarcoma is the most common (5, 6). Our study also found that most elderly patients with MBT over 60 years old are chondrosarcoma. Giuffrida et al. (7) found considerable differences in the 5-year survival rate of patients with different chondrosarcoma subtypes. The 5-year survival rate of dedifferentiated chondrosarcoma is 0%, while the exact cell type reaches 100%, the mucinous type is 71%, the paracortical type is 93%, the mesenchymal chondrosarcoma is 48%, and the malignant chondroblastoma is 85%. Strotman et al. (10) found that the 5-year overall survival rate of dedifferentiated chondrosarcoma is only 18%. The prognosis is worse for patients with axial bone, tumor size $>$8 cm, and lung cancer metastasis. A study by Bielack et al. (2) showed that age, location, and metastasis are the prognostic factors of patients with osteosarcoma. Previous studies have reported the predictive factors of osteosarcoma and predicted overall survival and cancer-specific survival (26, 27). Chord sarcoma is also a malignant bone tumor with high incidence in elderly patients. Previous studies have reported that chordoma mainly occurs in patients over 30 years old, and most of them are people over 60 years old (28). Chord sarcoma rarely metastasizes, but the 5-year survival rate will drop to about 50% (12). The incidence and prognostic factors of MBT in elderly patients have not yet been reported in the literature. Accurate prediction of patient survival is conducive to future treatment and follow-up. Therefore, we developed and validated nomograms to predict the survival of elderly patients with MBT.

In our study, we found that the increase in age will reduce the survival rate of elderly patients. Previous studies have also reported that age is a risk factor for bone tumors (27, 29). The reason may be that the increase in age leads to a decrease in the suppression of tumors by the immune system and an increase in the probability of comorbidities. In addition, our study also

TABLE 2 | Univariate and multivariate analyses of OS in the training set.

	Univariate			Multivariate		
	HR	**95% CI**	***P***	**HR**	**95% CI**	***P***
Age						
60–64	Reference			Reference		
65–69	1.06	0.74–1.52	0.76	1.19	0.91–1.56	0.20
70–74	0.87	0.59–1.28	0.48	1.13	0.85–1.50	0.41
75–79	1.27	0.84–1.94	0.26	2.06	1.53–2.76	<0.001
≥80	1.51	1.03–2.22	0.03	2.29	1.764–3.00	<0.001
Race						
White	Reference			Reference		
Black	1.25	0.74–2.13	0.4	0.98	0.69–1.41	0.93
Other	0.61	0.35–1.06	0.08	0.60	0.41–0.88	0.01
Sex						
Male	Reference			Reference		
Female	0.76	0.59–0.97	0.03	0.84	0.70–0.99	0.04
Year of diagnosis						
2004–2010	Reference					
2010–2018	0.56	0.43–0.72	<0.001			
Histologic type						
Chondrogenic sarcoma	Reference			Reference		
Osteogenic sarcoma	2.78	2.04–3.78	<0.001	1.25	0.98–1.60	0.08
Chordal sarcoma	0.7	0.49–0.99	0.04	0.61	0.434–0.86	0.01
Primary site						
Limb	Reference			Reference		
Cranial	0.56	0.37–0.84	<0.001	1.19	0.86–1.63	0.30
Spine	0.86	0.54–1.35	0.51	2.05	1.45–2.90	<0.001
Thoracic	0.32	0.21–0.5	<0.001	0.57	0.41–0.77	<0.001
Pelvic	0.97	0.7–1.32	0.83	0.98	0.77–1.26	0.89
Grade						
I	Reference			Reference		
II	1.77	1.02–3.06	0.04	1.19	0.84–1.69	0.34
III	5.39	3.1–9.37	<0.001	2.5	1.76–3.65	<0.001
IV	10.64	6.1–18.57	<0.001	2.88	1.969–4.22	<0.001
Unknown	2.14	1.26–3.64	<0.001	1.71	1.161–2.52	0.01
Stage						
Localized	Reference			Reference		
Regional	2.22	1.66–2.98	<0.001	1.61	1.31–1.98	<0.001
Distant	8	5.36–11.95	<0.001	1.79	1.17–2.72	0.007
T						
T1	Reference					
T2	2.95	2.27–3.83	<0.001			
T3	2.57	1.12–5.9	0.03			
T4	0	0–9.8	0.97			
N						
N0	Reference					
N1	1.77	0.79–3.93	0.16			
M						
M0	Reference			Reference		
M1	7.47	4.78–11.68	<0.001	1.623	1.04–2.53	0.03
Surgery						
No	Reference			Reference		

(Continued)

TABLE 2 | Continued

	Univariate			Multivariate		
	HR	**95% CI**	***P***	**HR**	**95% CI**	***P***
Partial resection	0.33	0.23–0.48	<0.001	0.429	0.33–0.56	<0.001
Radical resection	0.43	0.31–0.59	<0.001	0.397	0.31–0.51	<0.001
Amputation	0.71	0.46–1.11	0.14	0.36	0.26–0.50	<0.001
Radiation						
No/Unknown	Reference					
Yes	1.41	1.07–1.87	0.02			
Chemotherapy						
No/Unknown	Reference					
Yes	3.56	2.56–4.96	<0.001			
Tumor size	1.01	1.01–1.01	<0.001	1.00	1.00–1.01	<0.001

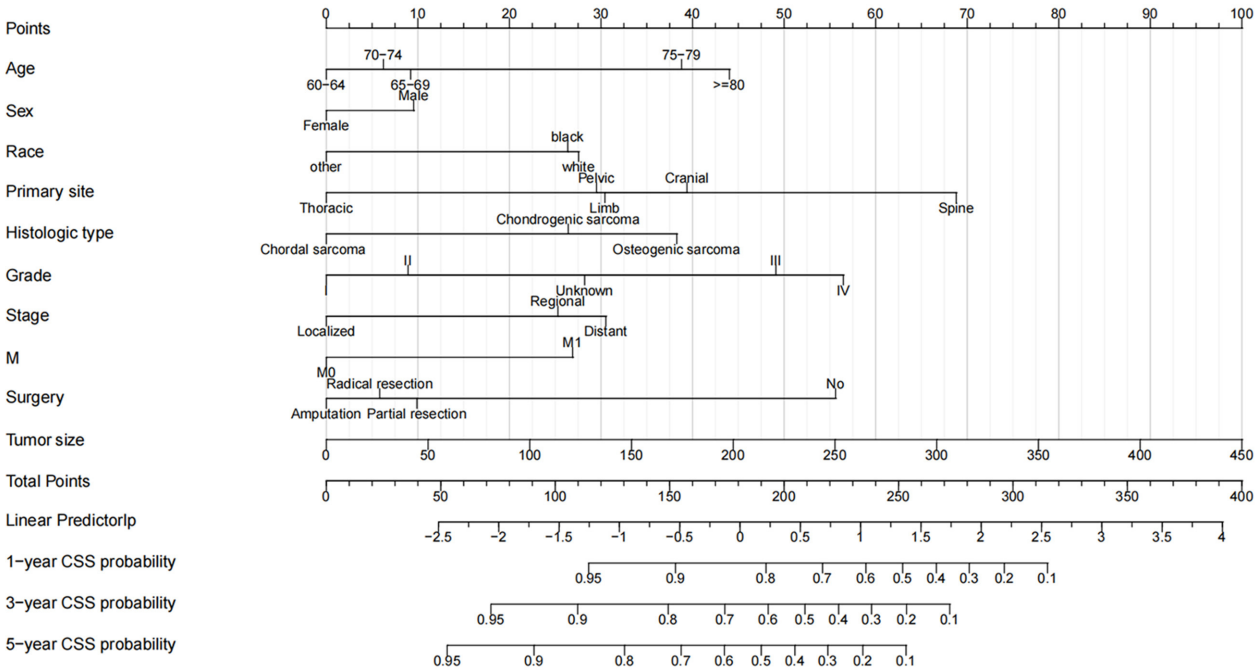


FIGURE 2 | Nomogram for 1-, 3-, and 5-year OS of elderly patients with MBT. The first row of the nomogram is the scoring ruler, rows 2–11 are the variables, and the 12th row is the patient's total score. 13–16 are the full score corresponding to the patient's 1-, 3-, and 5- survival rates.

found that the tumor site and size are also important factors affecting the prognosis. Similar to previous studies, patients with axial and large tumors have a higher risk of death (30–32). It may be because larger tumors are more likely to metastasize, and tumors in the axial position, especially the spine, are more likely to metastasize and invade surrounding tissues (33). Previous studies have found that the tumor stage is also a significant risk factor. The prognosis of distantly metastatic tumors is worse than localized tumors (34, 35). Our study also found that the tumor stage is an independent risk factor for prognosis. Besides, we found that sex and race are also significant risk factors, which have not been seen in previous studies. Women have a better survival prognosis than men. American Indians and Asians seem to have higher survival rates than whites and blacks.

In this study, we found that surgery significantly improved the prognosis of patients, similar to the results of previous studies (36, 37). In addition, according to risk stratification, we found that patients in the high-risk group benefited the most from partial tumor resection. At the same time, there is no significant difference in the prognosis of patients in the low-risk group with various treatment methods. Since most of the patients in the high-risk group are in the late stage of the disease and

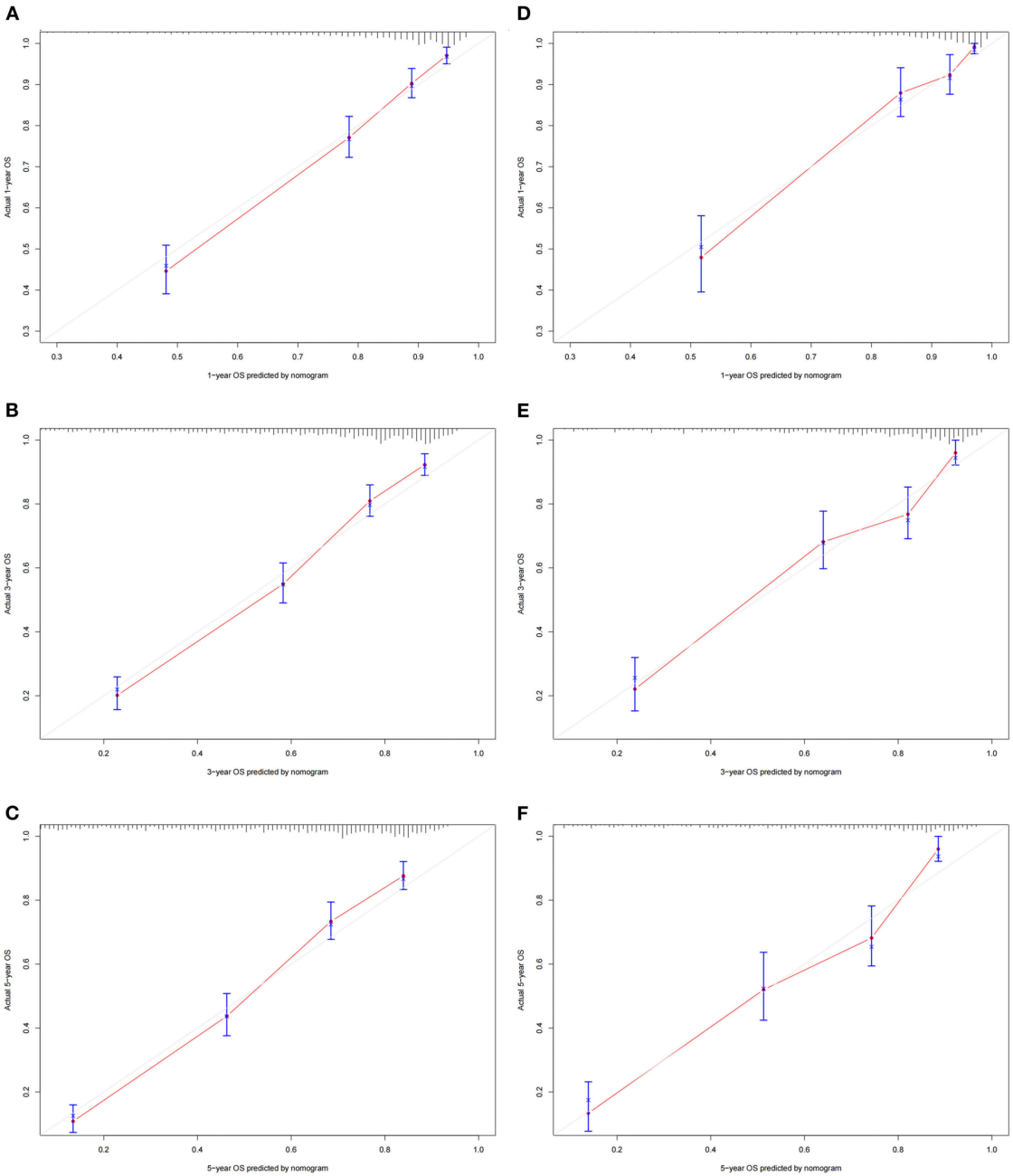


FIGURE 3 | Calibration curves of the nomogram. **(A–C)** For 1-, 3-, and 5-year OS in the training set; **(D–F)** For 1-, 3-, and 5-year OS in the validation set. The horizontal axis is the predicted value of the nomogram, and the vertical axis is the actual observed value of the nomogram. The coincidence between the predicted curve and the diagonal line means that the predicted value and the actual observed value are almost the same.

have already developed distant metastases, radical resection does not seem to improve the prognosis of the patients. This has some enlightenment for doctors and patients in choosing surgical methods for patients with different risks. On chemotherapy, our study found that chemotherapy does not affect patient survival, similar to previous studies (38).

Although our nomogram includes three histological types of MBT, it has been validated that the prediction model has good accuracy and discrimination. The C-index of the training set and the validation set were 0.779 [0.759–0.799] and 0.801 [0.772–0.830], respectively, proving the discriminative ability of the nomogram. The predicted value of the calibration curve is

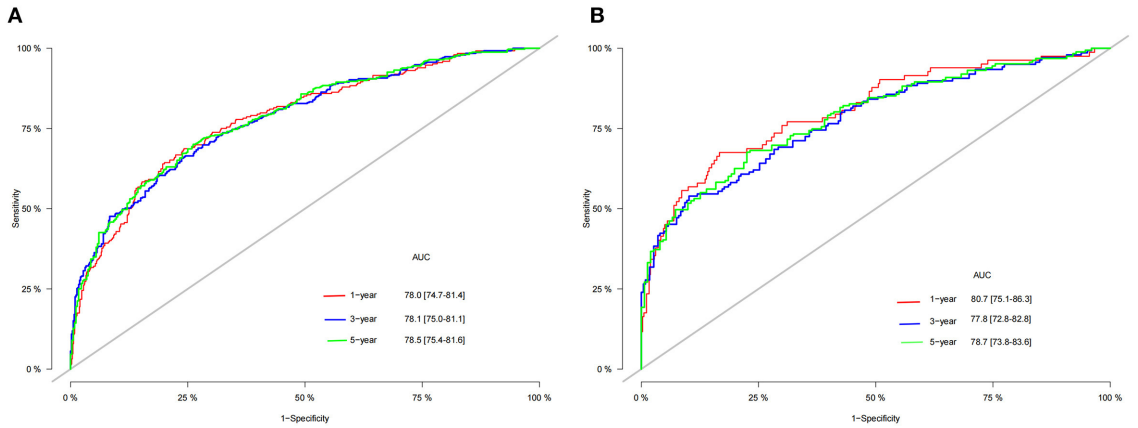


FIGURE 4 | The AUC of 1-, 3- and 5-year training set **(A)** and validation set **(B)**.

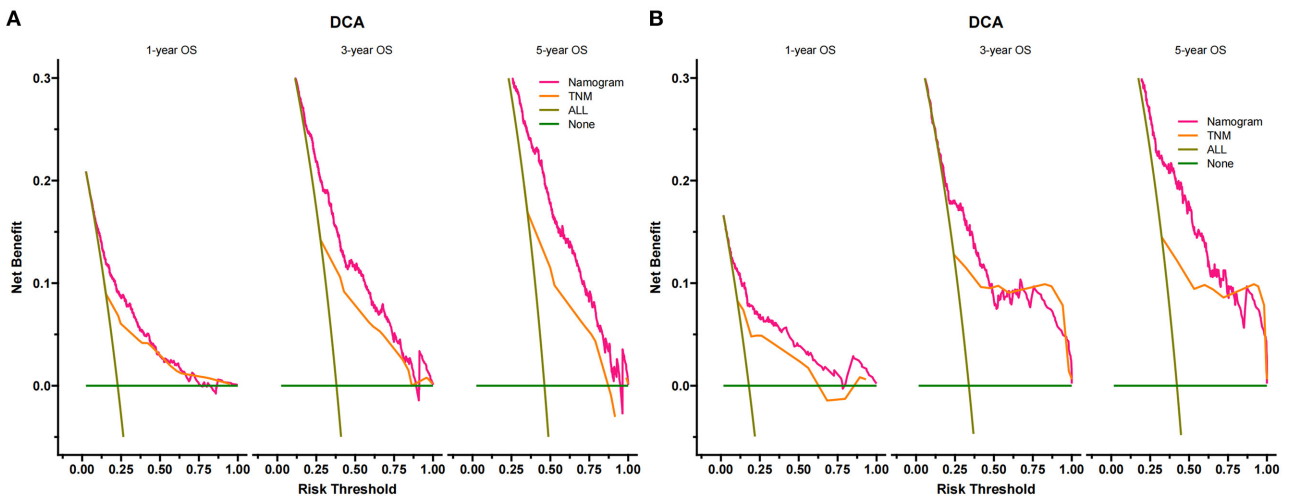


FIGURE 5 | Decision curves of the nomogram predicting OS in the training set **(A)** and validation set **(B)**. The y-axis represents the net benefit, and the x-axis represents the threshold probability. The green line indicates that no patients have died, and the dark green line indicates that all patients have died. When the threshold probability is between 20 and 100%, the net benefit of the model exceeds all deaths or no deaths.

highly consistent with the observed value, which indicates the accuracy of the prediction model. We compared the 1-, 3-, and 5-year DCA of the nomogram with the traditional TNM staging system and proved that the clinical value of the nomogram is higher than that of the traditional TNM staging system. Risk stratification also accurately distinguished high-risk and low-risk patients, and treatment and follow-up strategies should be different for patients in other risk groups. Patients in the high-risk group should choose appropriate treatment and receive closer follow-up.

This study is the first to explore risk factors for death in elderly patients with MBT using data from the SEER database. The overall survival rate of elderly patients with MBT was predicted based on risk factors. In other words, when elderly patients with MBT are admitted to hospitals, doctors can accurately inform patients of the 1-, 3- and 5-year survival rates based on these critical prognostic factors and make corresponding treatment decisions and follow-up strategies according to the survival rate of patients. This nomogram is of great significance for elderly patients with MBT.

However, our study still has some limitations. First, the SEER database cannot obtain detailed information on some variables such as surgical margins, BMI, smoking, and drinking, so it has a specific impact on prediction accuracy. However, we include essential variables such as tumor stage, size, surgery, and other factors that affect the prognosis of patients so that the results will not cause a devastating deviation. Second, The lack of information on some critical variables will cause errors in the prediction results. For example,

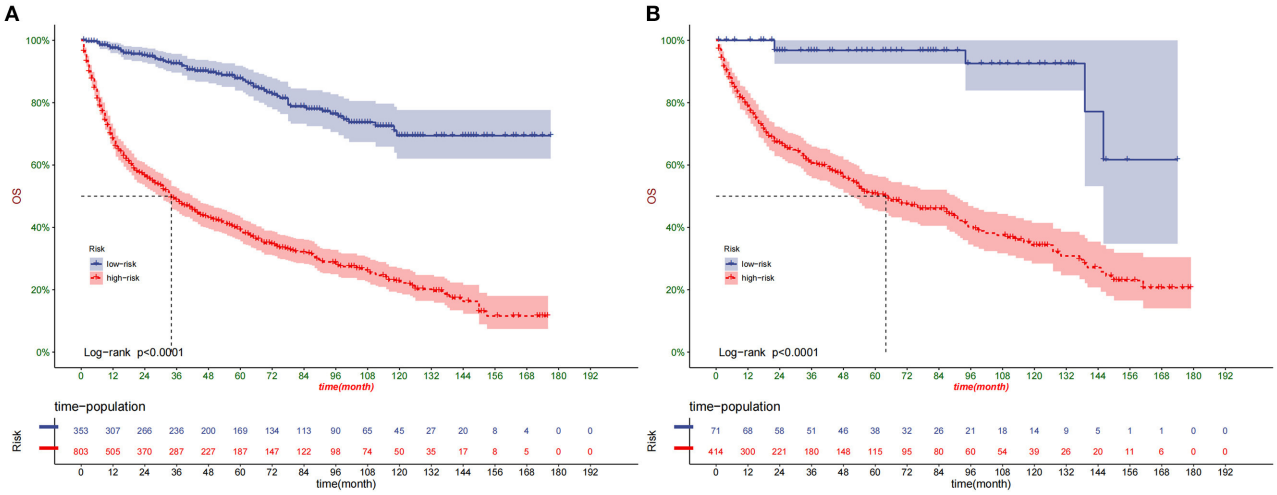


FIGURE 6 | Kaplan–Meier curves of OS for patients in the low- and high-risk groups in the training set **(A)** and validation set **(B)**.

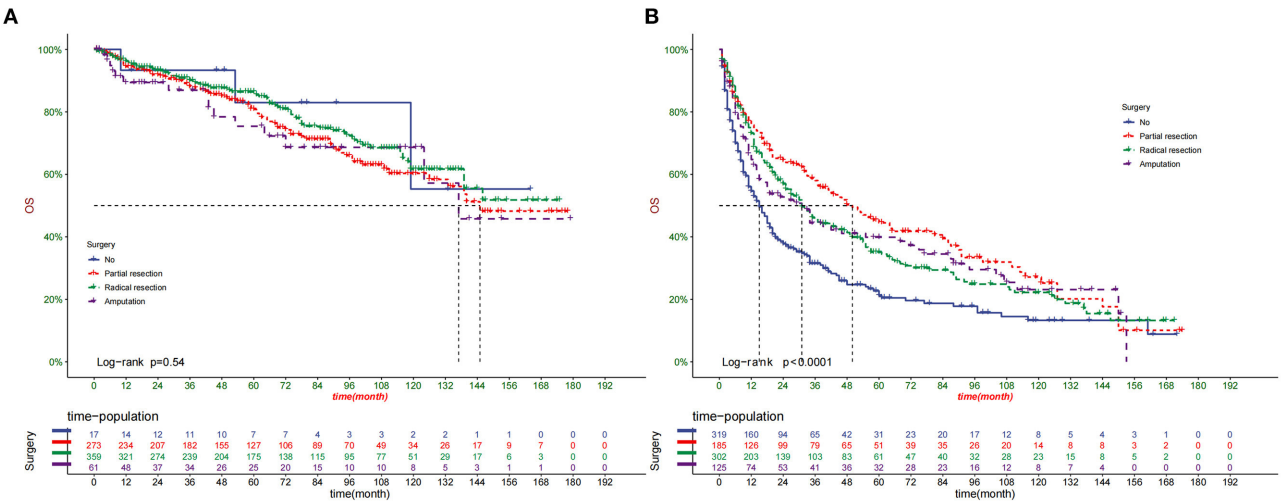


FIGURE 7 | Comparison of different surgical methods of Kaplan–Meier curves in low-risk **(A)** and high-risk **(B)** groups.

the tumor grade of some patients is unknown. However, it can be seen from the nomogram that the lack of tumor grade of some patients did not cause serious bias. Finally, our nomogram has only been validated internally, and external validation is necessary to test the predictive power of the nomogram.

CONCLUSION

A comprehensive nomogram was constructed to predict the 1-, 3-, and 5-year overall survival of elderly patients with MBT. This nomogram has been validated to have good accuracy and reliability and help patients and doctors predict survival prognosis and formulate treatment and follow-up strategies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov.

ETHICS STATEMENT

The public and anonymous data shared in this study is obtained from the SEER database.

AUTHOR CONTRIBUTIONS

JT, JW, and XP: contributed to the conception and design. JW and JT: collected, analyzed the data, drew the figures and tables, and wrote the draft. XP and JT contributed to manuscript writing and revision. All authors approved the final manuscript.

REFERENCES

1. He JP, Hao Y, Wang XL, Yang XJ, Shao JF, Guo FJ, et al. Review of the molecular pathogenesis of osteosarcoma. *Asian Pac J Cancer Prev.* (2014) 15:5967–76. doi: 10.7314/APJCP.2014.15.15.5967
2. Bielack SS, Carrle D, Hardes J, Schuck A, Paulussen M. Bone tumors in adolescents and young adults. *Curr Treat Options Oncol.* (2008) 9:67–80. doi: 10.1007/s11864-008-0057-1
3. Youn P, Milano MT, Constine LS, Travis LB. Long-term cause-specific mortality in survivors of adolescent and young adult bone and soft tissue sarcoma: a population-based study of 28,844 patients. *Cancer.* (2014) 120:2334–42. doi: 10.1002/cncr.28733
4. Franke M, Hardes J, Helmke K, Jundt G, Jürgens H, Kempf-Bielack B, et al. Solitary skeletal osteosarcoma recurrence. findings from the cooperative osteosarcoma study group. *Pediatr Blood Cancer.* (2011) 56:771–6. doi: 10.1002/pbc.22864
5. Li W, Dong S, Wang H, et al. Risk analysis of pulmonary metastasis of chondrosarcoma by establishing and validating a new clinical prediction model: a clinical study based on SEER database. *BMC Musculoskelet Disord.* (2021) 22:529. doi: 10.1186/s12891-021-04414-2
6. Saito R, Senbokuya N, Yagi T, Yoshioka H, Kanemaru K, Kinouchi H. Primary Spinal Intradural Extramedullary Mesenchymal Chondrosarcoma. *World Neurosurg.* (2021) 145:376–80. doi: 10.1016/j.wneu.2020.10.007
7. Giuffrida AY, Burgueno JE, Koniaris LG, Gutierrez JC, Duncan R, Scully SP. Chondrosarcoma in the United States (1973 to 2003): an analysis of 2890 cases from the SEER database. *J Bone Joint Surg Am.* (2009) 91:1063–72. doi: 10.2106/JBJS.H.00416
8. Damron TA, Ward WG, Stewart A. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: National Cancer Data Base Report. *Clin Orthop Relat Res.* (2007) 459:40–7. doi: 10.1097/BLO.0b013e318059b8c9
9. Schneiderman BA, Kliethermes SA, Nystrom LM. Survival in Mesenchymal Chondrosarcoma Varies Based on Age and Tumor Location: A Survival Analysis of the SEER Database. *Clin Orthop Relat Res.* (2017) 475:799–805. doi: 10.1007/s11999-016-4779-2
10. Strotman PK, Reif TJ, Kliethermes SA, Sandhu JK, Nystrom LM. Dedifferentiated chondrosarcoma: A survival analysis of 159 cases from the SEER database (2001-2011). *J Surg Oncol.* (2017) 116:252–7. doi: 10.1002/jso.24650
11. McMaster ML, Goldstein AM, Bromley CM, Ishibe N, Parry DM. Chordoma: incidence and survival patterns in the United States, 1973-1995. *Cancer Causes Control.* (2001) 12:1–11. doi: 10.1023/A:1008947301735
12. Bergh P, Kindblom LG, Gunterberg B, Remotti F, Ryd W, Meis-Kindblom JM. Prognostic factors in chordoma of the sacrum and mobile spine: a study of 39 patients. *Cancer.* (2000) 88:2122–34. doi: 10.1002/ (sici)1097-0142 (20000501)88:9<2122::aid-cncr19>3.0.co;2-1
13. Bacci G, Longhi A, Ferrari S, Briccoli A, Donati D, De Paolis M, et al. Prognostic significance of serum lactate dehydrogenase in osteosarcoma of the extremity: experience at Rizzoli on 1421 patients treated over the last 30 years. *Tumori.* (2004) 90:478–84. doi: 10.1177/030089160409000507
14. Whelan JS, Jinks RC, McTiernan A, Sydes MR, Hook JM, Trani L, et al. Survival from high-grade localised extremity osteosarcoma: combined results and prognostic factors from three European Osteosarcoma Intergroup randomised controlled trials. *Ann Oncol.* (2012) 23:1607–16. doi: 10.1093/annonc/mdr491
15. Bruns J, Elbracht M, Niggemeyer O. Chondrosarcoma of bone: an oncological and functional follow-up study. *Ann Oncol.* (2001) 12:859–64. doi: 10.1023/A:1011162118869
16. Lee FY, Mankin HJ, Fondren G, Gebhardt MC, Springfield DS, Rosenberg AE, et al. Chondrosarcoma of bone: an assessment of outcome. *J Bone Joint Surg Am.* (1999) 81:326–38. doi: 10.2106/00004623-199903000-00004
17. Fiorenza F, Abudu A, Grimer RJ, Carter SR, Tillman RM, Ayoub K, et al. Risk factors for survival and local control in chondrosarcoma of bone. *J Bone Joint Surg Br.* (2002) 84:93–9. doi: 10.1302/0301-620X.84B1.0840093
18. Hassan N, Slight R, Weiand D, Akke Vellinga G, Morgan F, Aboushareb S. Slight Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. *Int J Med Inform.* (2021). 150:104457. doi: 10.1016/j.ijmedinf.2021.104457
19. Alhazmi A, Alhazmi Y, Makrami A, et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med.* (2021) 50:444–50. doi: 10.1111/jop.13157
20. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain metastasis detection using machine learning: a systematic review and meta-analysis. *Neuro Oncol.* (2021) 23:214–25. doi: 10.1093/neuonc/noaa232
21. Yan Y, Liu H, Mao K, Zhang M, Zhou Q, Yu W, et al. Novel nomograms to predict lymph node metastasis and liver metastasis in patients with early colon carcinoma. *J Transl Med.* (2019) 17:193. doi: 10.1186/s12967-019-1940-1
22. Zhang Z, Xie H, Chen P, Cao P. Development and Identification of a Nomogram Prognostic Model for Patients with Primary Clear Cell Carcinoma of the Liver. *Med Sci Monit.* (2020) 26:e919789. doi: 10.12659/MSM.919789
23. Cronin KA, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. *Cancer.* (2014) 120 Suppl 23:3755–7.
24. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* (2008) 8:53. doi: 10.1186/1472-6947-8-53
25. Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin.* (2020) 70:7–30. doi: 10.3322/caac.21590
26. Zhang J, Yang J, Wang HQ, Pan Z, Yan X, Hu C, et al. Development and validation of a nomogram for osteosarcoma-specific survival: A population-based study. *Medicine.* (2019) 98:e15988. doi: 10.1097/MD.0000000000015988
27. Zheng W, Huang Y, Chen H, Wang N, Xiao W, Liang Y, et al. Nomogram application to predict overall and cancer-specific survival in osteosarcoma. *Cancer Manag Res.* (2018) 10:5439–50. doi: 10.2147/CMAR.S177945
28. Pan Y, Lu L, Chen J, Zhong Y, Dai Z. Analysis of prognostic factors for survival in patients with primary spinal chordoma using the SEER Registry from 1973 to (2014). *J Orthop Surg Res.* (2018) 13:76. doi: 10.1186/s13018-018-0784-3
29. Song K, Song J, Chen F, Lin K, Ma X, Jiang J. Prognostic nomograms for predicting overall and cancer-specific survival of high-grade osteosarcoma patients. *J Bone Oncol.* (2018) 13:106–13. doi: 10.1016/j.jbo.2018.09.012
30. Clark JC, Dass CR, Choong PF, A. review of clinical and molecular prognostic factors in osteosarcoma. *J Cancer Res Clin Oncol.* (2008) 134:281–97. doi: 10.1007/s00432-007-0330-x
31. Meltzer PS, Helman LJ. New Horizons in the Treatment of Osteosarcoma. *N Engl J Med.* (2021) 385:2066–76. doi: 10.1056/NEJMra2103423
32. Jeon DG, Song WS, Cho WH, Kong CB, Cho SH. Proximal tumor location and fluid-fluid levels on MRI predict resistance to chemotherapy in stage IIB osteosarcoma. *Clin Orthop Relat Res.* (2014) 472:1911–20. doi: 10.1007/s11999-014-3521-1
33. Duchman KR, Gao Y, Miller BJ. Prognostic factors for survival in patients with high-grade osteosarcoma using the Surveillance, Epidemiology, and End Results (SEER) Program database. *Cancer Epidemiol.* (2015) 39:593–9. doi: 10.1016/j.canep.2015.05.001
34. Janeway KA, Barkauskas DA, Krailo MD, Meyers PA, Schwartz CL, Ebb DH, et al. Outcome for adolescent and young adult patients with osteosarcoma: a report from the Children's Oncology Group. *Cancer.* (2012) 118:4597–605. doi: 10.1002/cncr.27414
35. Bielack SS, Kempf-Bielack B, Delling G, Exner GU, Flege S, Helmke K, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol.* (2002) 20:776–90. doi: 10.1200/JCO.2002.20.3.776
36. Whelan JS, Davis LE. Osteosarcoma, Chondrosarcoma, and Chordoma. *J Clin Oncol.* (2018) 36:188–93. doi: 10.1200/JCO.2017.75.1743
37. Mukherjee D, Chaichana KL, Parker SL, Gokaslan ZL, McGirt MJ. Association of surgical resection and survival in patients with malignant primary osseous spinal neoplasms from the Surveillance, Epidemiology, and End Results (SEER) database. *Eur Spine J.* (2013) 22:1375–82. doi: 10.1007/s00586-012-2621-4
38. Jiang Y, Wang T, Wei Z. Construction and Validation of Nomograms for Predicting the Prognosis of Juvenile Osteosarcoma: A Real-World Analysis in the SEER Database. *Technol Cancer Res Treat.* (2020) 19:1533033820947718. doi: 10.1177/1533033820947718

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tang, Wang and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fpubh.2021.806723

Global Trends and Hotspots in Research on Rehabilitation Robots: A Bibliometric Analysis From 2010 to 2020

Xiali Xue[1†], *Xinwei Yang*[2†], *Zhongyi Deng*[1], *Huan Tu*[1], *Dezhi Kong*[1], *Ning Li*[1*] and *Fan Xu*[3*]

[1] Institute of Sports Medicine and Health, Chengdu Sport University, Chengdu, China, [2] School of Sports Medicine and Health, Chengdu Sport University, Chengdu, China, [3] School of Public Health, Chengdu Medical College, Chengdu, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Mauro Callejas-Cuervo, Universidad Pedagógica y Tecnológica de Colombia, Colombia
Rajesh Kaluri, VIT University, India

***Correspondence:**
Ning Li
lining@cdsu.edu.cn
Fan Xu
xufan@cmc.edu.cn

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *01 November 2021*
Accepted: *02 December 2021*
Published: *11 January 2022*

Citation:
Xue X, Yang X, Deng Z, Tu H, Kong D, Li N and Xu F (2022) Global Trends and Hotspots in Research on Rehabilitation Robots: A Bibliometric Analysis From 2010 to 2020. Front. Public Health 9:806723. doi: 10.3389/fpubh.2021.806723

Background: In recent years, with the development of medical science and artificial intelligence, research on rehabilitation robots has gained more and more attention, for nearly 10 years in the Web of Science database by journal of rehabilitation robot-related research literature analysis, to parse and track rehabilitation robot research hotspot and front, and provide some guidance for future research.

Methods: This study employed computer retrieval of rehabilitation robot-related research published in the core data collection of the Web of Science database from 2010 to 2020, using CiteSpace 5.7 visualization software. The hotspots and frontiers of rehabilitation robot research are analyzed from the aspects of high-influence countries or regions, institutions, authors, high-frequency keywords, and emergent words.

Results: A total of 3,194 articles were included. In recent years, the research on rehabilitation robots has been continuously hot, and the annual publication of relevant literature has shown a trend of steady growth. The United States ranked first with 819 papers, and China ranked second with 603 papers. Northwestern University ranked first with 161 publications. R. Riener, a professor at the University of Zurich, Switzerland, ranked as the first author with 48 articles. The *Journal of Neural Engineering and Rehabilitation* has the most published research, with 211 publications. In the past 10 years, research has focused on intelligent control, task analysis, and the learning, performance, and reliability of rehabilitation robots to realize the natural and precise interaction between humans and machines. Research on neural rehabilitation robots, brain–computer interface, virtual reality, flexible wearables, task analysis, and exoskeletons has attracted more and more attention.

Conclusions: At present, the brain–computer interface, virtual reality, flexible wearables, task analysis, and exoskeleton rehabilitation robots are the research trends and hotspots. Future research should focus on the application of machine learning (ML), dimensionality reduction, and feature engineering technologies in the research and development of rehabilitation robots to improve the speed and accuracy of algorithms. To achieve

wide application and commercialization, future rehabilitation robots should also develop toward mass production and low cost. We should pay attention to the functional needs of patients, strengthen multidisciplinary communication and cooperation, and promote rehabilitation robots to better serve the rehabilitation medical field.

Keywords: machine learning, bibliometric analysis, CiteSpace, trend, artificial intelligence, rehabilitation robot

INTRODUCTION

In the era of the COVID-19 pandemic, reducing doctor–patient contact is an important measure to prevent crossinfection. The use of rehabilitation robots to assist healthcare workers in treating patients with COVID-19 has reduced doctor–patient contact and increased productivity. Rehabilitation robotics is a relatively young and rapidly growing field with increasing penetration into the clinical environment (1, 2). Functional rehabilitation and the auxiliary robot have gradually become important technical means of clinical rehabilitation treatment in the world and have spawned several new rehabilitation robot technologies and systems (3). They have received great attention from all countries and have developed with each passing day (4). Rehabilitation robots, originating from the research field of engineering, are the perfect combination of rehabilitation medicine and robot technology, which makes up for the deficiency of traditional rehabilitation treatment methods in ensuring the high intensity of rehabilitation training, the persistence of endurance, and the standardization of training effects. They integrate the knowledge of artificial intelligence, biomechanics, information science, and rehabilitation medicine and use intelligent bionic technology to assist patients to complete limb-training movements and achieve the purpose of rehabilitation. These robots are used mainly for the assistance and rehabilitation of the daily life of the disabled (5–8).

According to the functional classification, rehabilitation robots are mainly divided into three types: full-body rehabilitation robots (9), upper limb robots (10), and lower limb rehabilitation robots (11). According to the means of application, they can be divided into fixed wearable robots, such as exoskeleton robots. The exoskeleton robot is mainly used to assist walking and is a special type of rehabilitation robot. Stationary robots are mainly used for rehabilitation training. Exoskeletons and soft-bodied robots are divided according to how the driving force is transmitted (12–14). Functional rehabilitation robots are mainly applied to chronic patients with stroke, cerebral palsy, spinal cord injury, paraplegia, traumatic brain injury, and limb injury. Due to the limited function recovery effect of traditional rehabilitation training methods on chronic injury, rehabilitation robot-assisted training is being studied and applied more and more (15–20). At present, rehabilitation robots have been widely used in rehabilitation nursing, prosthesis, and rehabilitation therapy, which not only promotes the development of rehabilitation medicine but also drives the development of new technologies and new theories in related fields. At the same time, there are still many challenges in the development of rehabilitation robots, there are still controversies in the research on rehabilitation robots, such as safety problems, and the actual operation is not simple and convenient, so that it is limited to large hospitals; moreover, the number of people who are suitable and can use rehabilitation robots is small, and the actual utilization rate is low (21, 22). For various rehabilitation robots with different characteristics and advantages, the application effect of different rehabilitation programs is not clear, so it is particularly important to formulate the user guide of rehabilitation robots to guide the clinical application of rehabilitation robots.

CiteSpace visual analysis software is a Java application that generates cocitation networks based on reference citations to reveal the structure of a particular research area, enabling visual exploration through knowledge discovery in bibliographic databases. Through progressive knowledge domain visualization, the most cited and critical documents, areas of expertise within the knowledge domain, and the emergence of research topics can be intuitively plotted to detect and visualize trends and patterns in the scientific literature (23). The key issue in evidence-based medical practice is to find the best available evidence about a clinical problem, and the task can be simplified by combining visual analysis techniques with traditional methods developed in evidence-based medicine (24–27). After consulting the database, it was found that the research in the field of rehabilitation robots increased significantly after 2010. Therefore, this study took the literature related to rehabilitation robots collected in the core data collection of the Web of Science database from 2010 to 2020 as the research object (28). It uses CiteSpace 5.7 visualization software, combining the qualitative analysis of the online platform and the traditional method of literature metrology, research focus from the literature field, high-influence countries or regions, institutions, the author, the high-frequency keywords and mutation term, etc., on the international research hotspot in the field of rehabilitation robots and frontier parsing and tracking. The study also provides some references for the research on rehabilitation robots and the development of related subjects. At the same time, the future development trend of rehabilitation robots is prospected, which is of great significance to improve the function of rehabilitation robots and to develop and improve the future medical field.

METHODS

Data Sources and Retrieval Strategies

The core data set of the Web of Science database was retrieved by computer, and the retrieval condition was "TS = (Rehabilitation Robot OR Rehabilitation Robotics)." The period is set as December 31, 2010 to December 31, 2020. The article is a peer-reviewed journal paper with good authority and representativeness. To ensure high-quality visualization results, the literature type in this study is limited to "article," and the

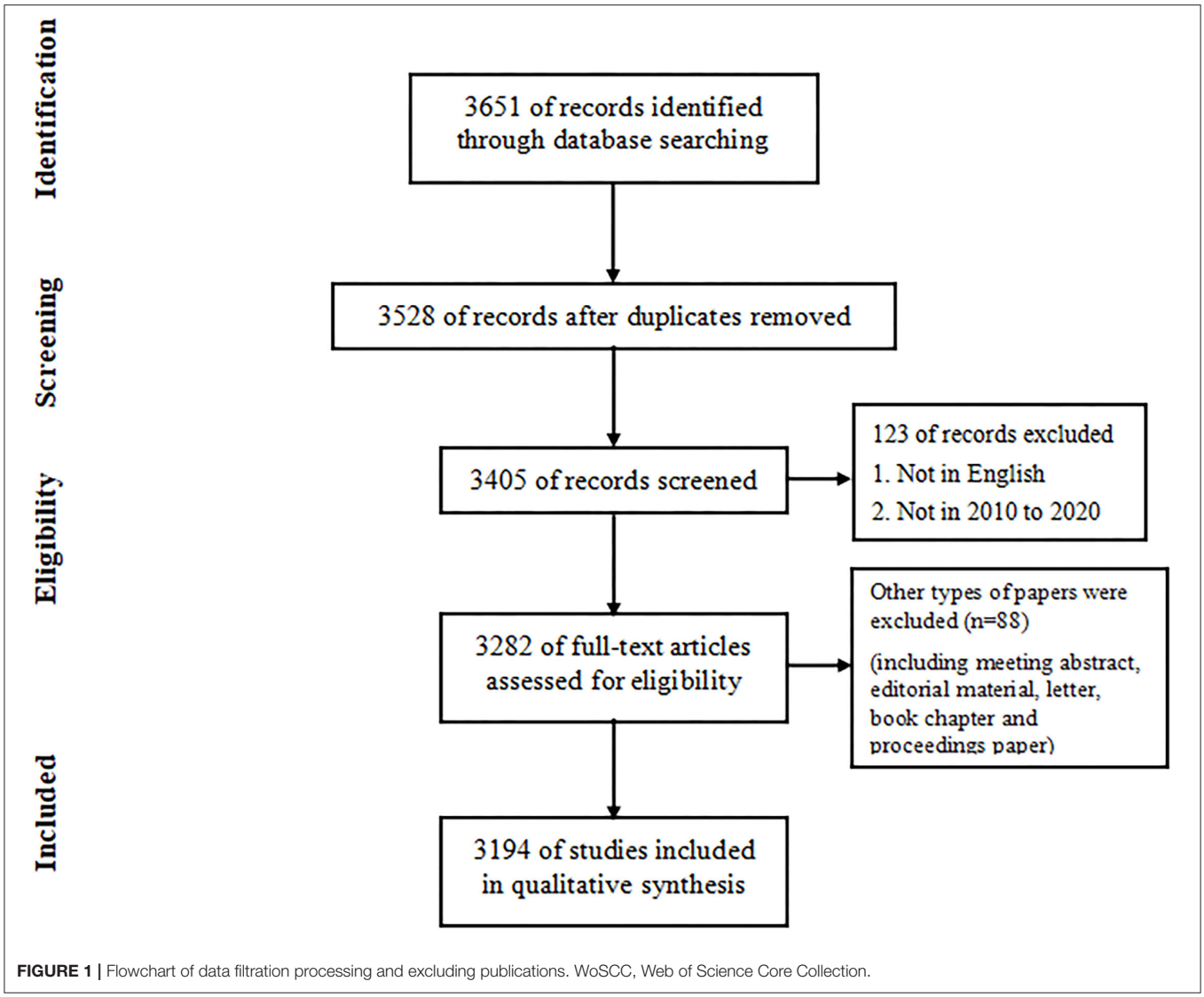


FIGURE 1 | Flowchart of data filtration processing and excluding publications. WoSCC, Web of Science Core Collection.

language type is set as "English." It excludes conference abstracts, letters, reviews, news, journal reviews, and other literary types.

Literature Screening

Two evaluators read the literature independently. First, preliminary screening was carried out according to the title and abstract of the article, and then screening was carried out again according to the inclusion and exclusion criteria. After the screening, if there was any dispute, the third evaluator read the whole paper, discussed it together, and decided whether or not to reject it.

Data Analysis

CiteSpace 5.7 visualization software determines the information related to the literature according to the information of the first author in the article, mainly including the country, author, keywords, research institution, and so on. In the CiteSpace 5.7 visual software parameter setting, the period is set to 2010–2020; the time scale is set to 1 year; the threshold is set to "top N" and set to 50; and the top 50 representative papers are selected for systematic review and analysis. We chose Pathfinder as the clipping connection to simplify the network structure and highlight important features. Countries (regions), institutions, authors, cited authors, and keywords were selected for node type cooccurrence analysis to draw a visual atlas. CiteSpace was used to detect the mutation rate of keywords and generate a keyword ranking table with a high mutation rate. In the network graph, different nodes represent various elements, such as authors, institutions, countries, and keywords, whereas the size of nodes reflects the number or frequency of publications, and the connections between nodes represent relations such as cooperation, cooccurrence, or cocitation (23, 29–31).

RESULTS

Bibliometric of Publication Output

A total of 3,194 publications met the inclusion criteria from 2010 to 2020 (**Figure 1**). **Figure 2** shows the total number

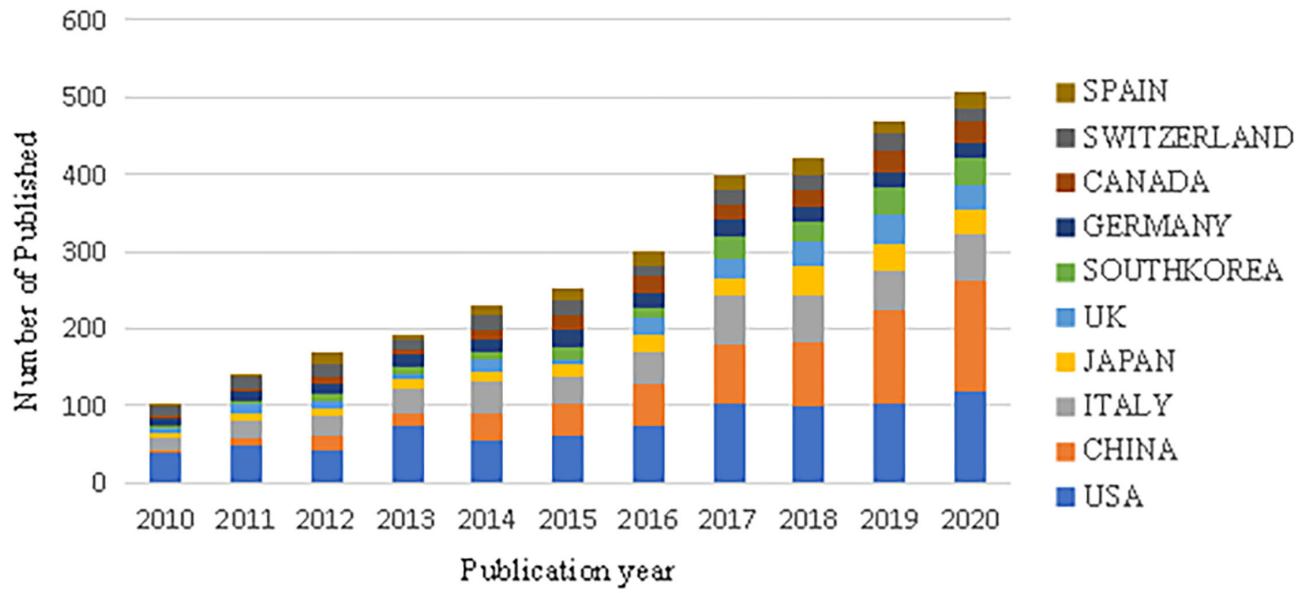


FIGURE 2 | Bibliometric analysis of WoS core database output. From 2010 to 2020, the number of published publications by different countries on rehabilitation robot research has changed over the years.

of publications annually, where the number of papers in rehabilitation robot research maintained a gradually increasing trend from 2010 to 2020.

Hot Topics in Literature Research

Among the 3,194 retrieved publications, according to the distribution analysis of literature research hotspots, the literature with more than 50 articles was classified into 25 categories (**Figure 3**). Of these, 809 belong to the category of rehabilitation medicine, 721 biomedical engineering, and 589 neuroscience. It can be seen from the data that rehabilitation robots are widely studied and applied in the fields of rehabilitation medicine, biomedical engineering, and neuroscience.

Country or Geographic Analysis

A total of 82 countries or regions contributed to the research on rehabilitation robots, as shown in **Table 1**, among which the United States published the most articles (819). China ranked second with 603 publications. The third was Italy, with 444. A total of 2,605 institutions have published research related to rehabilitation robotics, the most prolific of which is Northwestern University (USA), with 161 publications. The second is the University of Tsukuba (Japan), with 133 publications. The University of Maryland ranked third with 81 publications (**Table 1**). Visualization of the study countries shows that the top three countries in centrality are Spain (0.4), the USA (0.26), and Germany (0.17). These three countries have established good research cooperation with other countries. The detailed cooperation between countries or regions is shown in **Figure 4**.

Research Hotspots Based on Cooccurrence of Research Institutions

Through the analysis of cooperative cooccurrence of research institutions, CiteSpace software was used to draw the view of research hotspots of cooperative cooccurrence of research institutions of rehabilitation robots. A total of 411 nodes and 616 connections were generated, and the density of the topological network was 0.0073. Visualization of research institutions shows that Scuola Super Sant Anna (0.24), Northwestern University (0.23), and MIT (0.12) rank in the top three in centrality, respectively. These three institutions have established good research cooperation with other institutions (**Figure 5**).

High-Influence Authors and Cooperative Relationships

R. Riener, a professor at the University of Zurich in Switzerland, is the most prolific author with 48 articles, followed by S.K. Agrawal, a professor at Columbia University in the US, and Y. Sankai, a professor at the University of Tsukuba in Japan, with 29 articles. G. Kwakkel, a professor at the University of Amsterdam (512 times), followed by S. Hesse, a professor at Freie Universitat Berlin (512 times), and J. Mehrholz, a professor at Technische Universitat Dresden (335 times) (**Table 2**). In the author collaboration atlas, Professor R. Riener holds the leading

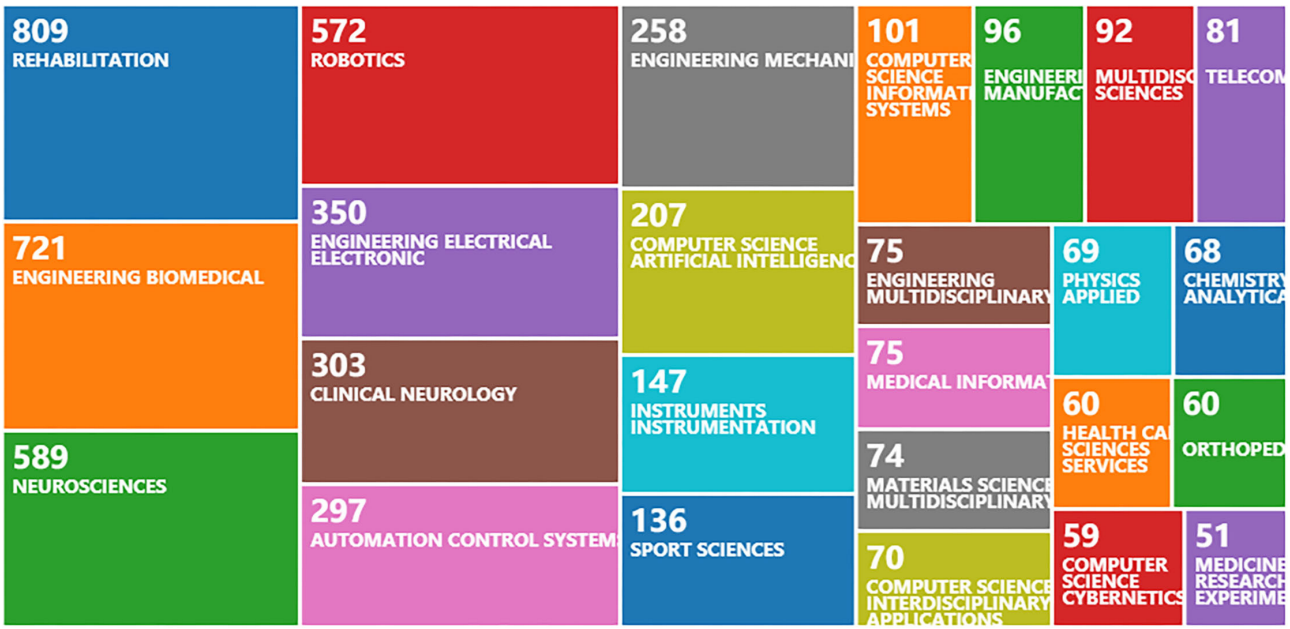


FIGURE 3 | 2010–2020 rehabilitation robot literature research hotspot distribution.

TABLE 1 | Influence of high-yield countries and institutions on rehabilitation robot research literature from 2010 to 2020.

Ranking	Countries/regions	No. of publications	Centrality	Institutions	No. of publications
1	USA	819	0.26	Northwestern Univ	161
2	China	603	0.08	Univ Tsukuba	133
3	Italy	444	0.15	Univ Maryland	81
4	Japan	221	0.06	Scuola Super Sant Anna	81
5	UK	208	0.15	Univ Auckland	80
6	South Korea	187	0.03	Chang Gung Univ	80
7	Germany	181	0.17	Natl Univ Singapore	79
8	Canada	180	0.15	MIT	76
9	Switzerland	177	0.07	Columbia Univ	75
10	Spain	159	0.4	Univ Twente	61

position in this field by absolute advantage. However, there is no cooperative relationship with other authors with a high number of publications, and the cooperative relationship between other highly influential authors is also average (**Figure 6**).

High-Influence Journals

Among the 3,194 articles retrieved, there are 515 source journals and the top 10 journals in terms of publication volume (**Table 3**). The top three journals in terms of publications are J Neuroeng Rehabil, IEEE T Neur Sys Reh, and IEEE Robot Autom Let.

Highly Cited Literature

Among the 3,194 retrieved articles, the total citation frequency was 47,224, and the top 10 cited (32–41) kinds of literature with high citation frequency (**Table 4**). The top three sources are LANCET, NEJM, and ROBOT AUTON SYST.

Research Hotspots Based on Keyword Cooccurrence

Through keyword cooccurrence analysis, 20 high-frequency keywords with a frequency greater than 150 times were statistically ranked. The top three high-frequency keywords were "rehabilitation" (1,276 times), "stroke" (940 times), and "design" (471 times). The top three high-frequency keywords in centrality are "robot" (0.15), "walking" (0.15), and "movement" (0.15) (**Table 5**). CiteSpace software was used to draw the hotspot view of rehabilitation robot research. A total of 607 nodes and 2,728 connections were generated, and the density of the topological network was 0.0148 (**Figure 7**).

Keywords Clustering Research Hotspots and Frontier

CiteSpace was used for clustering analysis of keywords, and the classic LLR algorithm was adopted to obtain nine clustering

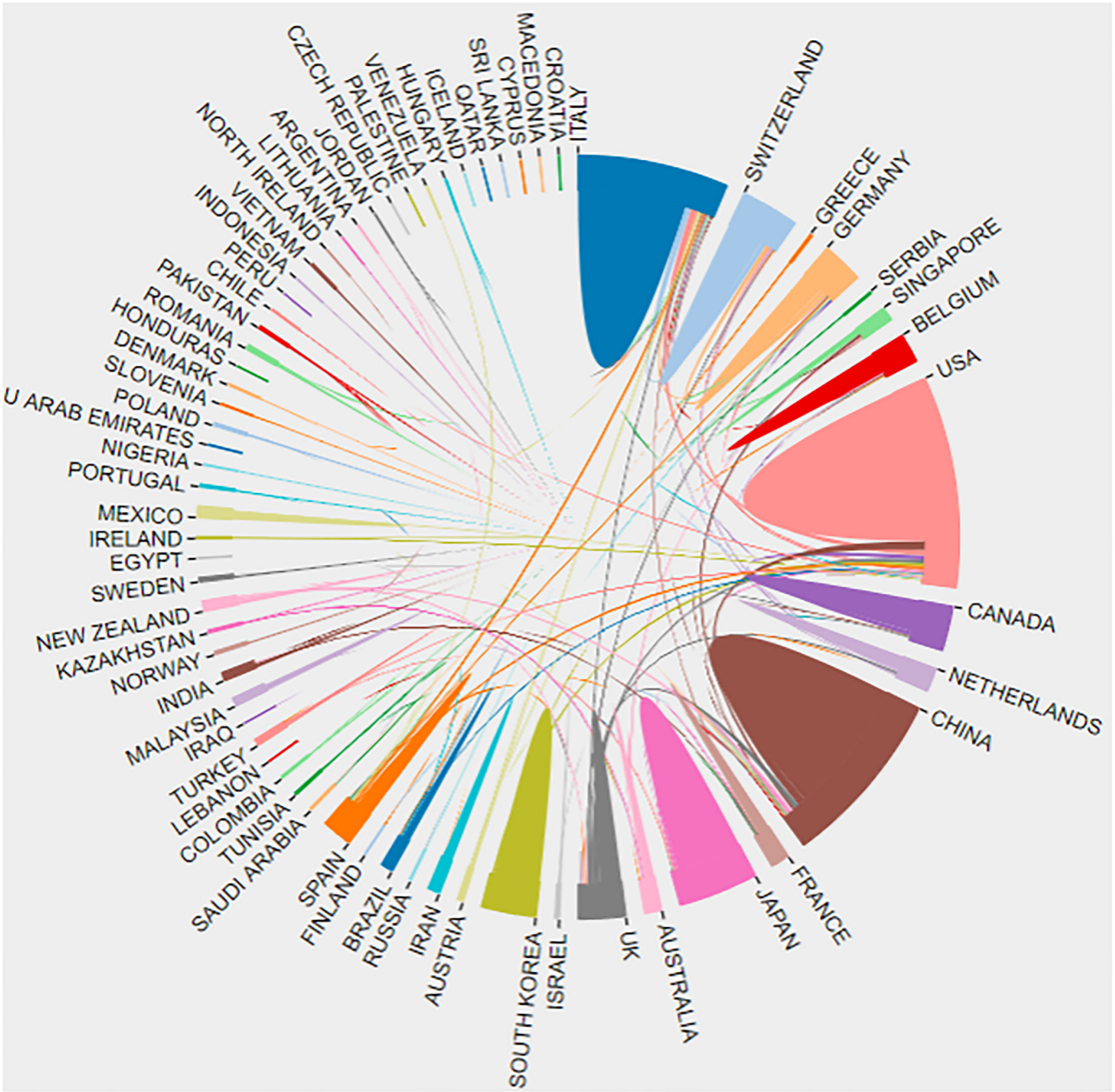


FIGURE 4 | Bibliometric analysis of the cooperation of countries in the field of rehabilitation robot research.

groups: (0) walking, (1) stroke, (2) brain–computer interface, (3) stroke, (4) electromyography, (5) proprioception, (6) task analysis, (7) soft actuators, and (8) soft robotics (**Figure 8**). The timeline view is used to display the dynamic time change of clustering keywords. From 2010 rehabilitation robots, cerebral apoplexy, exoskeleton, the stability of the robot, rehabilitation, spinal cord injury, stroke, nerve rehabilitation, motor learning, virtual reality, and other keywords widely attention, among them, the rehabilitation robot and cerebrovascular accident, exoskeleton robot hand, adaptive control, and touch the heart of the study on the relationship between the recovery began in 2011. Rehabilitation robot and brain–computer interface, upper limb rehabilitation, and meta-analysis research began to appear in 2014. Research on the relationship between rehabilitation robots and upper limb rehabilitation and meta-analysis was hot until 2019. In 2019–2020, robot sensing system, actuator, pneumatic artificial muscle, telemedicine, task analysis, wearable robot, and flexible robot were new words. It is predicted that further research will focus on rehabilitation robots and robot sensing systems, pneumatic artificial muscle, wearable robot, and flexible

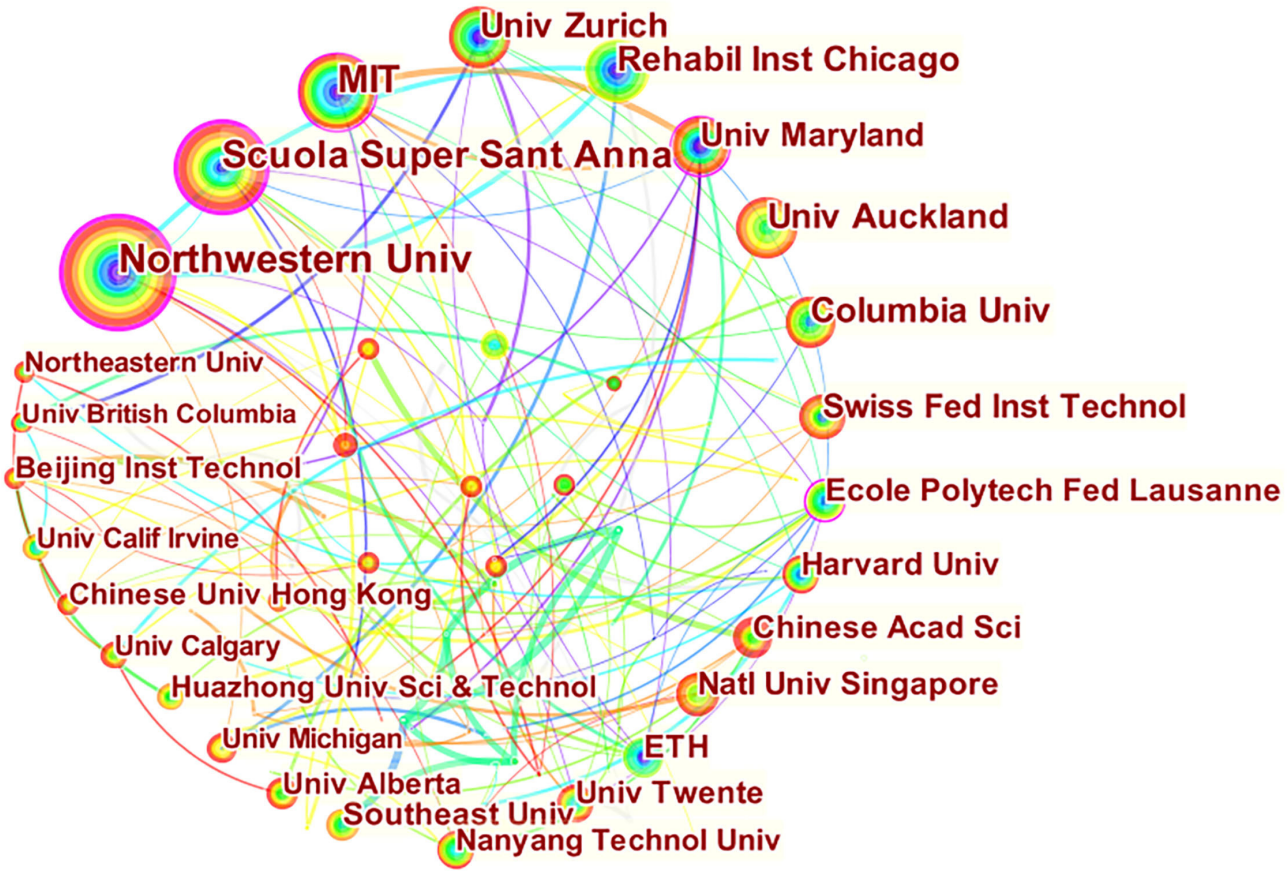


FIGURE 5 | Co-occurrence atlas of rehabilitation robot research institutions. The bigger the circle, the more it appears. Lines refer to the connections between institutions. The thicker the lines, the closer the connections.

TABLE 2 | Status of authors with high influence on rehabilitation robot research from 2010 to 2020.

Ranking	Highly prolific authors	Country	No. of publications	Highly cited authors	Total cites	Country
1	Riener, R	Switzerland	48	Kwakkel, G	512	Netherlands
2	Agrawal, SK	USA	33	Hesse, S	438	Germany
3	Sankai, Y	Japan	29	Mehrholz, J	426	Germany
4	Kadone, H	Japan	23	Krebs, HI	393	USA
5	Yamazaki, M	Japan	23	Hogan, N	320	USA
6	Calabro, RS	Italy	23	Riener, R	308	Switzerland
7	Naro, A	Italy	21	Lo, AC	306	USA
8	Song A	China	18	Lum, PS	302	USA
9	Hada, Y	Japan	18	Veneman, JF	275	Spain
10	Micera, S	Switzerland	17	Reinkensmeyer, DJ	268	USA

robot. The time dynamic evolution of keywords is presented in **Figure 9**.

Cocitation Clustering Analysis of Related Kinds of Literature

Through literature cocitation cluster analysis, CiteSpace software was used to draw a cocitation map of the rehabilitation robot research literature, generating 868 nodes and 1,884 connections, and the density of the topological network was 0.005. With the lowest citation times of 15, 28 of the 3,194 cited kinds of literature were included from the research scope, and the top three cited kinds of literature with high centrality were "Lo et al. (33)" (0.88), "Maciejasz et al. (42)" (0.68), and "Klamroth-Marganska et al. (39)" (0.45).

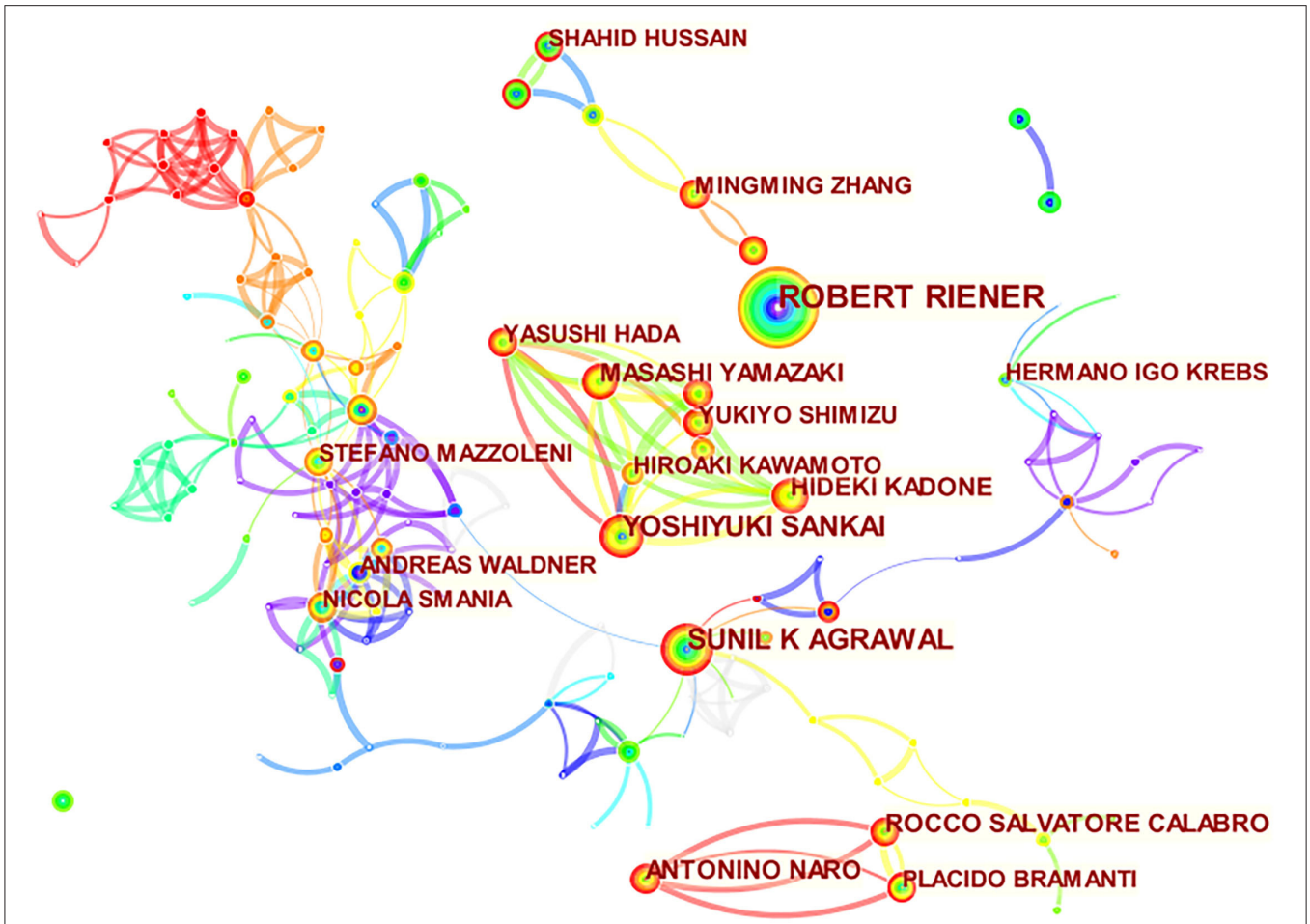


FIGURE 6 | Author and cooperative relationship. The bigger the circle, the more it appears. Lines refer to the connections between authors. The thicker the lines, the closer the connections.

TABLE 3 | Status of high-impact journals on rehabilitation robot research from 2010 to 2020.

Ranking	Journals	IF (2020)	JCR (2020)	No. of publications	Total cites	Average total cites
1	J Neuroeng Rehabil	4.262	Q1	211	1,174	5.56
2	IEEE T Neur Sys Reh	3.800	Q1	167	821	4.92
3	IEEE Robot Autom Let	3.743	Q2	69	110	1.59
4	Sensors	3.574	Q1	66	65	0.98
5	IEEE Access	3.745	Q2	64	36	0.56
6	IEEE-ASME T Mech	5.673	Q1	59	488	8.27
7	Neurorehabilitation	2.134	Q2	55	200	3.64
8	Neurorehab Neural RE	3.914	Q1	54	603	11.17
9	Front Neurorobotics	2.650	Q2	51	72	1.41
10	Arch Phys Med Rehab	3.964	Q1	48	217	4.52

Funding Support Institutions

There are 2,637 funded institutions, among which the top 10 funded institutions have funded 1,535 studies, accounting for 48% of the literature in the field of rehabilitation robot research. The top three funders were the China National Natural Science Research Foundation (337), the US Department of Health and Human Services (247), and the US National Institutes of Health (240). These results indicate that China has strong support for the research on rehabilitation robots and a large amount of research fund investment, which may be related to the incentive

TABLE 4 | References with high citation frequency of rehabilitation robots from 2010 to 2020.

Ranking	Title	Publication year	Total cites	Journals	IF (2020)	JCR (2020)
1	Stroke Care 2 stroke rehabilitation (20)	2011	952	Lancet	79.320	Q1
2	Robot-assisted therapy for long-term upper-limb impairment after stroke (21)	2010	704	NEJM	91.240	Q1
3	Soft robotic glove for combined assistance and at-home rehabilitation (22)	2014	485	Robot Auton Syst	3.123	Q1
4	The ReWalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury (23)	2012	346	Am J Phys Med Rehab	2.152	Q2
5	Getting neurorehabilitation right: what can be learned from animal models? (24)	2012	281	Neurorehab Neural RE	3.914	Q1
6	Tactile-direction-sensitive and stretchable electronic skins based on human-skin-inspired interlocked microstructures (25)	2014	279	ACS Nano	15.884	Q1
7	An EMG-based control for an upper-limb power-assist exoskeleton robot (26)	2012	249	IEEE T Syst Man Cy B	13.450	Q1
8	Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomized trial (27)	2014	237	Lancet Neurol	44.182	Q1
9	Design and control of a bio-inspired soft wearable robotic device for ankle-foot rehabilitation (28)	2014	213	Bioinspir Biomim	2.952	Q1
10	Current hand exoskeleton technologies for rehabilitation and assistive engineering (29)	2012	212	Int J Precis Eng Man	2.104	Q2

TABLE 5 | High-frequency keywords of rehabilitation robot research from 2010 to 2020 (frequency > 150).

Ranking	Keywords	Frequency	Centrality	Ranking	Keywords	Frequency	Centrality
1	Rehabilitation	1,276	0.14	11	Gait	309	0.09
2	Stroke	940	0.13	12	System	289	0.06
3	Design	471	0.1	13	Arm	274	0.06
4	Recovery	469	0.12	14	Upper limb	256	0.11
5	Exoskeleton	434	0.11	15	Movement	238	0.15
6	Robot	418	0.15	16	Spinal cord injury	217	0.08
7	Therapy	373	0.1	17	Performance	198	0.15
8	Walking	345	0.15	18	Reliability	189	0.08
9	Rehabilitation robotics	326	0.11	19	Impairment	167	0.09
10	Robotics	316	0.06	20	Upper extremity	163	0.05

mechanism initiated by the government. However, there is still a gap between the total number and that of the United States, which needs to be improved (**Table 6**).

Frontier Analysis of an Emergent Word Detection Algorithm

We used CiteSpace to detect emergent words. A total of 108 mutants were detected. Among the 25 keywords with the highest citation frequency, the mutation cycles of 16 mutation words were concentrated from 2010 to 2015. The strongest mutation intensity was "arm" (8), followed by "motor control" (7.13). The third place was "hemiparetic patient" (6.13), the fourth place was "body weight support" (5.14), and the fifth place was "arm movement" (5.11). Keywords with high mutation intensity in the last 5 years include "modulation" (2015–2016), "computer interface" (2015–2017), "treadmill therapy" (2016–2017), "series elastic actuator" (2017–2020), and "prosthetics and exoskeleton" "wearable robotics" (2018–2020). These emergent words in the past 5 years may also be the main direction of future research.

DISCUSSION

International Research Status of Rehabilitation Robots

The research on rehabilitation robots can be traced back to the 1960s, but the development was relatively slow due to

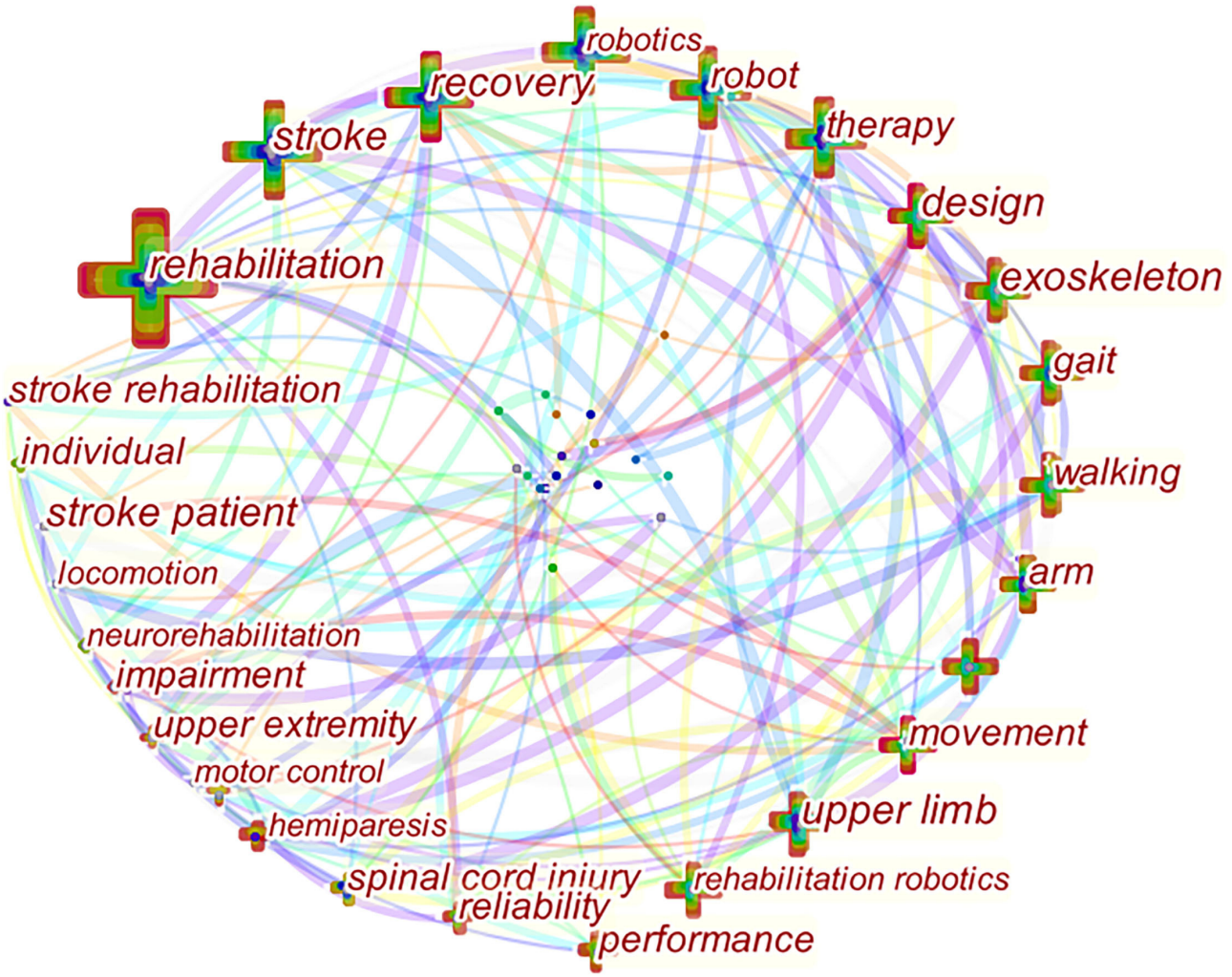


FIGURE 7 | Co-occurrence spectrum of high-frequency keywords in rehabilitation robot field. The cross represents keywords; the larger the cross, the thicker the cross, the higher the frequency of keywords; the line refers to the connection between keywords; the thicker the line, the closer the connection.

design and funding problems. The 1980s was the initial stage of the research on rehabilitation robots, and the United States and the United Kingdom were at the leading level. After the 1990s, the research on rehabilitation robots entered a period of comprehensive development (43). In 1987, the British man Michael developed the prototype of the Handy I rehabilitation robot. In 1989, the Massachusetts Institute of Technology developed the first clinical trial robot for rehabilitation treatment. In 2020, Elon Musk reported the latest research results on the brain–computer interface: a coin-sized chip implanted into a pig's brain can read the pig brain's information in real time. A functional device implanted into a surgical robot can perform all the steps of craniotomy, implant sensors, and glue. With the continuous development of interdisciplinary integration, researchers are making progress in the field of rehabilitation robots. It can be predicted that the vision of human–machine integration will be realized soon (44, 45).

Research Hotspots by Country (Region), Author, and Founding

In this study, the research literature related to rehabilitation robots was obtained from the core data collection of the Web of Science database in the past 10 years as the research object, and CiteSpace visual analysis software was used to analyze and track the hotspots and frontiers of international research in the field of rehabilitation robots. This study presents the international research status of rehabilitation robots in the past 10 years intuitively and forms a general understanding of its research hotspots, frontiers, and trends. Overall, international rehabilitation robot research identified steady growth year by year, nerve rehabilitation, stroke, brain–machine interface, virtual reality, flexible wearables, task analysis, and the exoskeletons of rehabilitation robot research more and more aroused people's interests, becoming increasingly important areas of research.

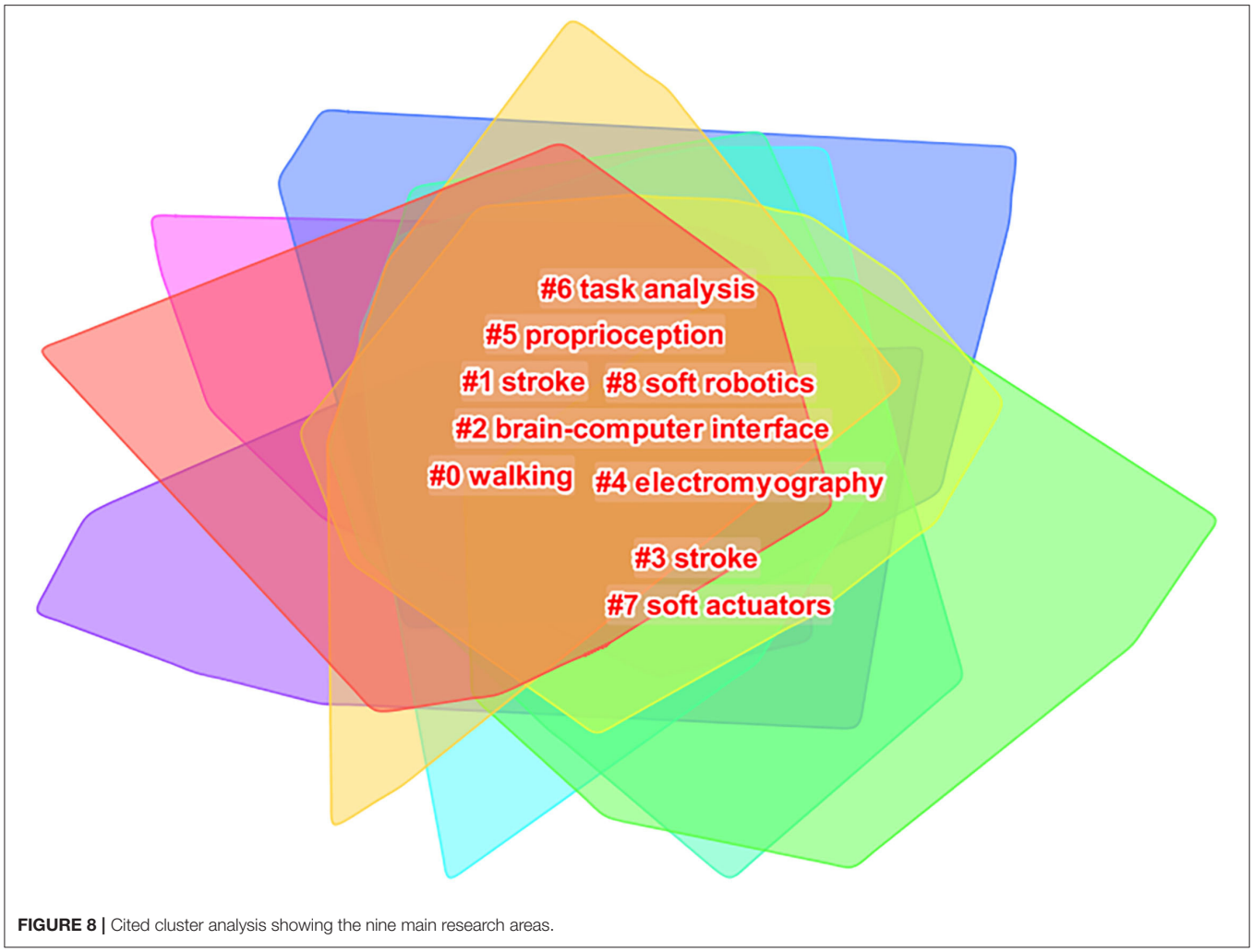


FIGURE 8 | Cited cluster analysis showing the nine main research areas.

The number of articles published in the world in the field of rehabilitation robots generally shows an upward trend, mainly focusing on the fields of rehabilitation medicine, biomedical engineering, and neuroscience, indicating that the research on rehabilitation robots is popular in the field of medicine. The United States, China, Italy, Japan, and the United Kingdom are the major research exporters. The United States has the most articles about rehabilitation robots, followed by China. China is the only developing country, which shows great progress in the field of rehabilitation robots in the past decade and playing an important role. In terms of research institutions, Northwestern University, the University of Tsukuba, the University of Maryland, Santa Anna University of Pisa, Italy, and the University of Auckland, New Zealand occupy a certain position in the field of rehabilitation robots with a high number of publications. Northwestern University ranked first with 161 publications.

Among the prolific authors, R. Riener, a professor at the University of Zurich in Switzerland, ranked first with 48 articles, and the second and third were from the United States and Japan, respectively. It is worth noting that Professor Song from Southeast University of China ranked eighth with 18 articles. Studies by Chinese scholars in the field of rehabilitation robots are moving toward the international community, even catching up abroad.

In terms of cocitation of authors, the top five are Prof G. Kwakkel, Vrije Universitat Amsterdam; Prof S. Hesse, Medical University of Charlotte, Berlin, Germany; Prof J. Mehrholz, Dresden Medical School, Technical University of Dresden, Germany; Prof H.I. Krebs, MIT, USA; and Prof N. Hogan, MIT, USA. It can be seen that scholars from the Netherlands, Germany, and the United States have high influence in the research on rehabilitation robots. The above data can provide potential collaborators and teams' information for future research. According to Price's law, half of all papers on the same subject are written by a group of highly productive authors roughly equal to the square root of the total number of authors (46). A total of 9,163 authors were included in 3,194 publications. Of these, the first 93 authors completed 50% of the papers, in line with Price's law. It indicates that the core authors in the field of rehabilitation robots have been established. Paying attention to these authors can help us better understand the frontier and trend

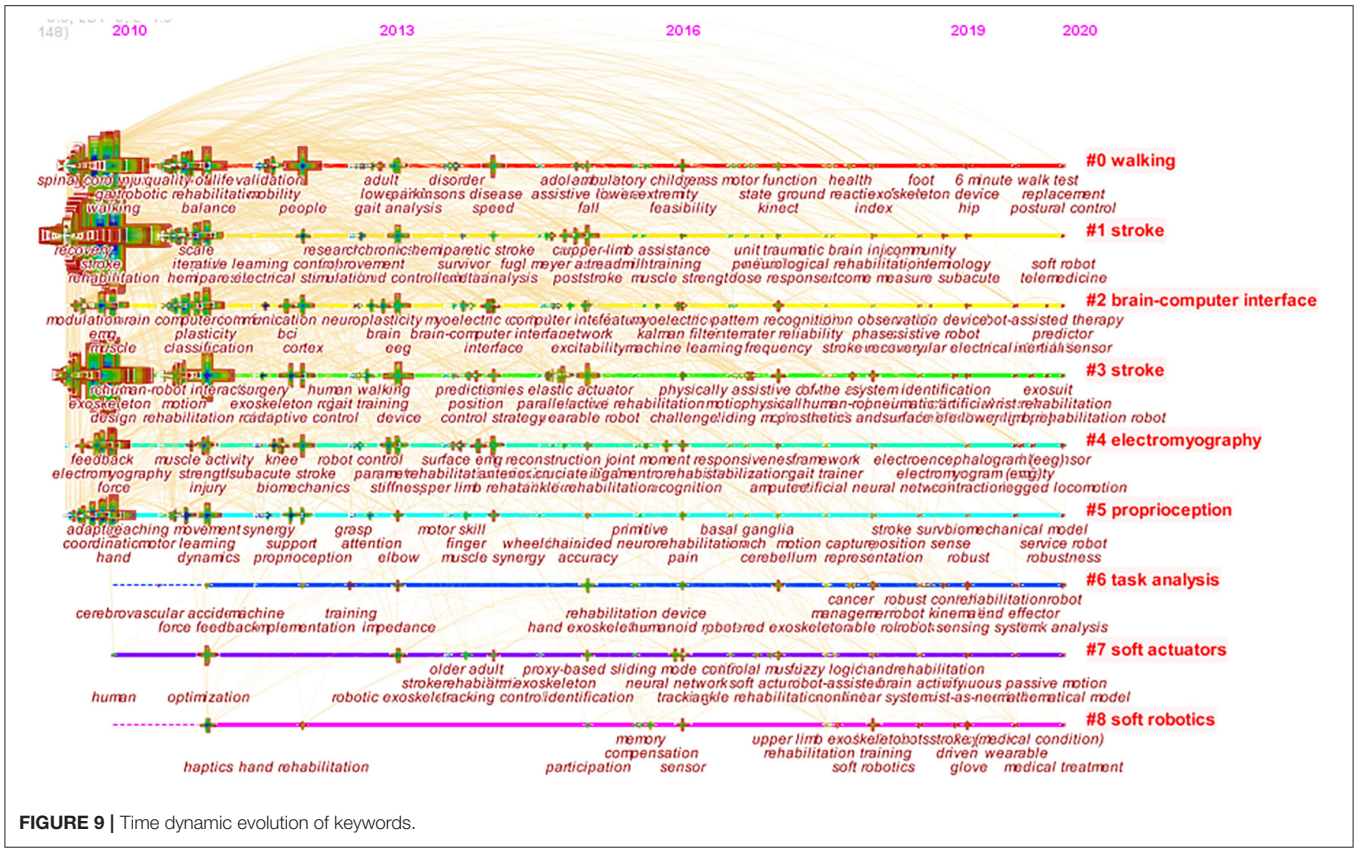


FIGURE 9 | Time dynamic evolution of keywords.

TABLE 6 | Funding institutions of rehabilitation Robot Fund from 2010 to 2020.

Ranking	Funding	No. of publications	Percentage
1	National Natural Science Foundation of China NSFC	337	10.541
2	United States Department of Health Human Services	247	7.726
3	National Institutes of Health NIH USA	240	7.507
4	National Science Foundation NSF	162	5.067
5	NIH Eunice Kennedy Shriver National Institute of Child Health Human Development NICHD	117	3.66
6	European Union EU	111	3.472
7	Ministry of Education Culture Sports Science and Technology Japan MEXT	95	2.972
8	Japan Society for the Promotion of Science	86	2.69
9	Grants in AID for Scientific Research Kakenhi	72	2.252
10	NIH National Institute of Neurological Disorders Stroke NINDS	68	2.127

of rehabilitation robots. Among the top five foundations, one is from China, and four are from the United States. The research area involves many scientific research grants, but the main funding areas are natural sciences, medicine, bioengineering, and material science.

Research Hotspots in the Field of Rehabilitation Robots

Citation status is one of the important indicators to reflect research hotspots. Through reading and analyzing the literature with high citation frequency and centrality, some research results with high attention in the research field of rehabilitation robots can be learned to reveal the areas of concern (47). Through reading and analysis of the top ten cited studies, the results show that stroke, spinal cord injury, paraplegia, upper limb functional rehabilitation, neuroplasticity, actuator technology, flexible robot gloves, etc., have received extensive attention from researchers in a certain stage in the research field of rehabilitation robots. The research hotspots of these highly cited papers help to identify the major hotspots of rehabilitation robot research. At present, the number of cited papers in China is low, which may be due to the late publication of many studies in other countries.

Keywords with high centrality and frequency represent research hotspots in a period. According to the analysis of the

hot keywords, research in the field of rehabilitation robots mainly focuses on rehabilitation, stroke, design, exoskeleton robotics, upper limbs, gait, spinal cord injury, performance, and reliability. Through keyword clustering analysis, the higher the degree of aggregation, the better the homogeneity between studies (48).

In the last 10 years, the rehabilitation of upper limb function for stroke has been a research hotspot. With the rapid growth of new technologies and equipment, there is much different training equipment for rehabilitation robots designed to move or assist arm movement. There are significant differences in the types of treatment provided by different devices in terms of the techniques employed and the treatments provided. In a systematic review, there is increasing evidence that rehabilitation robot training can benefit the recovery of arm function after stroke (49). Human–robot interaction (HRI) is one of the hot topics in the current research on rehabilitation robots, which is how to realize the natural and precise interaction between humans and machines through the fusion of machine intelligence and biological intelligence.

Research Frontiers and Trends in Rehabilitation Robots

Emergent words represent frequently cited keywords in a given period, thus indicating cutting-edge features and trends. Combined with the time dynamic evolution and burst analysis of keywords, we can get a general understanding of the research field, development, and future trend of rehabilitation robots. The top five post-2014 breakout words are "modulation," "computer interface," "treadmill therapy," "series elastic actuator," and "prosthetics and exoskeleton." Modulation refers to the adjustment and control of the signal and program of the rehabilitation robot to achieve the purpose of precise rehabilitation. Computer interface refers to the popular brain–computer interface technology. Treadmill therapy mainly refers to lower limb rehabilitation robots that are similar to treatments done on a treadmill. A Lokomat lower limb rehabilitation robot is a typical hanging weight loss standing rehabilitation robot of the treadmill therapy type, which combines with the treadmill platform to complete the exercise (50). In the control system, the series elastic actuator can convert control signals into power devices that can drive the controlled system, which is the core of the power development and design of the rehabilitation robot (51). Research on prosthetics and exoskeleton robots is currently a hot topic in comparison, and because of the rapid development of mechanical design and control algorithms for electro-mechanical systems, exoskeletons have been significantly developed but are still limited to larger body areas, such as the upper and lower limbs. Due to their small size and strong tactile perception, hand exoskeletons still face many challenges in many technical fields, such as hand biomechanics, neurophysiology, rehabilitation, actuators and sensors, human–robot physical interaction, and ergonomics.

In addition, machine learning (ML) is one of the technologies that have developed rapidly in the past 10 years. It is widely used in computer vision, bioinformatics, health care, business analysis, trend forecasting, and other fields. ML allows computers to learn from large samples of data and predict patterns that exist in the data (52, 53). ML algorithms to help realize the future of improved health care by unleashing the potential of large biomedical and patient data sets are used in different research areas to predict and classify accurate results from test data (54). They have proven to be helpful for diagnosis in a variety of medical and clinical data sets. In the field of rehabilitation, ML can help predict the recovery function of walking level of the patients with stroke and assist in research on the development of walking assistance robots. AI-based virtual reality rehabilitation (55) and forecasting mortality of patients with stroke after complete rehabilitation with tree-based ML models have been studied (56). For any type of human disease prediction, data sets need to be preprocessed. In this regard, dimensionality reduction plays a vital role in reducing the high-dimensional data into reduced dimensionality (57). Over the past few decades, several dimensionality reduction techniques have been used to filter data samples from the data set under consideration. Dimensionality reduction requires the mapping of higher-dimensional inputs to lower-dimensional inputs so that similar points in the input space are mapped to adjacent points on the manifold (54, 58). Feature engineering is an important preprocessing step that helps in the extraction of transformed features from the raw data that will simplify the ML model and also improve the quality of the results of an ML algorithm.

In the field of medical rehabilitation, more and more people tend to interact with machines naturally and accurately. With personalized AI task analysis, ML, electromechanical interaction, intelligent control, and other technologies combined with robots applied to the field of rehabilitation medicine, the rehabilitation robot "therapist" will become the main force in the field of rehabilitation therapy in future. The wearable exoskeleton robot can help elderly people with severe limb dysfunction, hemiplegia, and spinal cord injuries walk normally. In addition, it can help humans maintain balance function and increase strength to achieve "human–machine integration" (5). The "mind control" of brain–computer interface is gradually switching from science fiction to reality, helping the movement, feeling, and other functions of damaged limbs return to normal. In 2020, *MIT Technology Review* published a breakthrough in the brain–computer interface by Professor Eduardo Fernandez of Miguel Hernandez de Elche University: the electronic artificial eye, which restores sight to patients who are completely blind (59, 60).

With human intelligence and machine intelligence, combining human–computer interaction as a representative of technological breakthroughs between man and machine combined with the increasingly close, with the help of human–computer interaction technology and method, and combining human intelligence and machine intelligence make the two complementary advantages, to work together, and in the aspect of medical rehabilitation which spawned a significant theoretical innovation and breakthrough technology method. The brain–computer interface will be widely applied in future (61, 62). How to establish the ability of high-broadband data connection between human and machine is also an important trend in the current international rehabilitation robot research. At present, rehabilitation robots of limb function still cannot provide intuitive tactile neurofeedback,

and patients can only rely on visual feedback to judge the size of the object to be grasped and other sensory information. How to achieve accurate neural feedback of sensory information still needs further research. It is the challenge and trend of the development of rehabilitation robots in future to make that rehabilitation robots have emotional characteristics and make patients feel happy while helping patients to achieve their rehabilitation goals.

Regarding rehabilitation robot technology and the research and development and application of the system, there is still a gap between China and developed countries such as Europe and the United States. To improve the performance of rehabilitation robot system, it is necessary to master the key core technologies of rehabilitation robot, such as EMG information perception and control technology, force feedback control technology, joint angle and torque control technology, and space motion detection technology (4). In particular, the research on these technologies in brain–computer interface, virtual reality, robot sensing system, actuators, pneumatic artificial muscle, and the wearable flexible robot can better promote the development of rehabilitation robots to meet the future demand of Chinese people for health services.

Application and Development Trend of Rehabilitation Robots

It can be predicted that rehabilitation robots will be used more and more in an era when the cost of manual treatment is becoming more and more expensive. To better popularize and commercialize the application, future development needs to increase the universality of rehabilitation robots, focusing on lightweight, portable, reconfigurable, intelligent, and AI-based equipment, thus bringing new treatment methods and thinking to the rehabilitation field under the premise of ensuring safety, mass production, and cheap direction development (63). Intelligent rehabilitation is the future development trend (64). It should break through the barrier of the human–machine interface, develop multidisciplinary cooperation and communication, strengthen collaboration with rehabilitation medicine and artificial intelligence, actively participate in the design and development of rehabilitation robots, and promote rehabilitation robots to better serve the rehabilitation medicine field.

When thinking about the application and development of rehabilitation robots, Professor Li first mentioned the "human nature" of rehabilitation robots. Rehabilitation robots should have the basic human nature—namely, perception, thinking, expression and communication, action and cooperation, and learning and adaptability—and complete training and operations independently or with the assistance of an automatic machine (14). At present, rehabilitation robots cannot complete all the techniques and operations of professional rehabilitation personnel; particularly, the self-learning ability, adaptability, and creativity of robots need theoretical breakthroughs. The most difficult thing for rehabilitation robots to break through is the benevolence and moral sense of medical workers (65). The twenty first century is an era of health and life science. Health is the most powerful driving force for scientific and technological revolution and social development. Rehabilitation robots as one of the most important technologies in the human health industry will undoubtedly provide innovative means and paths for rehabilitation medicine, preventive medicine, and clinical medicine, making the previously invisible visible and the impossible possible (66). Rehabilitation robots will also provide strong impetus for the reform of the medical system, the renewal of the medical model, and the extension of the human healthy life span (67).

Study Limitations

This study included only studies on rehabilitation robots published in the core data collection of SCIE, the Web of Science database, and the paper type was English literature, which may have ignored high-quality kinds of literature in other languages in the field of rehabilitation robots, resulting in certain limitations in literature retrieval. In cluster analysis, different categories of content may overlap. In addition, the cocitation frequency of kinds of literature is related to time. The kinds of literature published in recent years may have a relatively low total cocitation frequency due to their short publication time, resulting in differences between research results and the actual situation. The cooccurrence of keywords is limited by the number of nodes, which may affect the accuracy of conclusions. However, accuracy itself is relative and does not affect the reliability of conclusions obtained in the paper. Due to different algorithms, there is no standardized setting process for time partition, threshold, or clipping mode in the process of using CiteSpace to generate visual atlas, which may cause some bias.

CONCLUSION

Based on the Web of Science database and CiteSpace software, this study for the first time conducted bibliometric and visual analysis of the international research on rehabilitation robots in the past 10 years from multiple perspectives and presented the research overview of rehabilitation robots in the past 10 years relatively scientifically and intuitively. At present, the brain–computer interface, virtual reality, flexible wearables, task analysis, and exoskeleton rehabilitation robots are the frontiers and hotspots of research, representing the development trend of future research, and they can be used as a reference direction for future research. Future research in the field of rehabilitation robots should focus on the following aspects: first, focus on the functional needs of patients to improve the quality of life; second, strengthen interdisciplinary cooperation and build a regional cooperation network among countries.

Challenges

Formulating better rehabilitation robots in such an evolving field is important. Until today, numerous technical challenges and insufficient safety standards have prevented the large-scale development of rehabilitation robots. As it is multidisciplinary technical research, it faces many challenges in the actual research and development process. The primary challenge of rehabilitation robots is safety, which is the most basic

requirement to help patients with rehabilitation training. On the premise of ensuring the effectiveness of the product, it is more important to ensure the safety of use and avoid secondary injuries to patients. Intelligent control, ergonomics, at present, and the movement energy consumption of the robot bionic man are much higher than that of the simple skeletal muscle movement, the operation process of the wearable exoskeleton robot is generally more complex, and the large weight makes the rehabilitation training more difficult, but the precision is much worse. The man–machine interface design is a challenging problem. This interface is the medium and dialogue window for transmitting and exchanging information between humans and computers. A good man–machine interface is key to improve the operation efficiency of rehabilitation robots. For example, patients' subjective feelings and physiological reactions, especially visual perception, cannot be timely perceived in the training process. Both need constant optimization and learning. Existing rehabilitation robots generally lack tactile feedback. It is suggested to study various sensors to enhance the interaction between rehabilitation robots and the human body, involving multidisciplinary cooperation and communication. In addition, a key challenge regarding the control of rehabilitation robots is to design shared control or ML technologies that are predictable to patients and do not exceed their needs. Because not all technical components of rehabilitation robots are well developed or designed for daily life and outdoor applications, a lot of collaborative work and utilization of resources from medical technology, biomechanics, engineering, and product development are needed. The price of the current rehabilitation robots is relatively expensive, which is a great burden on ordinary families. It is suggested that future research can find cost-effective materials to popularize rehabilitation robots in ordinary life.

Future Direction

In terms of technology, ML, dimensionality reduction, feature engineering, and other technologies play an important role in the research and development of rehabilitation robots. In future, the effectiveness of these technologies can be tested on high-dimensional data such as images and text data. They can also be used for more complex algorithms, such as deep neural networks, convolutional neural networks, and recursive neural networks. The new ideas of mixing and matching hardware, programs, algorithms, and expected neural pathways may lead to more focused research and, ultimately, significant advances in robot-assisted rehabilitation. For various rehabilitation robots with different characteristics and advantages, the application effect of different rehabilitation programs is not clear, so it is particularly important to formulate the user guide of rehabilitation robots to direct the clinical application of rehabilitation robots. To achieve widespread application and commercialization, future rehabilitation robots should be developed toward mass production and low cost. In addition, in future, rehabilitation robots will still be the focus and hotspot in the fields of rehabilitation medicine and intelligent engineering, and with the development of rehabilitation robots with different functions, their usage guidelines will become more and more perfect, and the cooperation between rehabilitation therapists and rehabilitation robots is bounded to open a new chapter in rehabilitation therapy.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: when used on a reasonable request, it can be obtained by contacting the corresponding author. Requests to access these datasets should be directed to lining@cdsu.edu.cn.

AUTHOR CONTRIBUTIONS

XX and XY: designing this study and writing initial draft and revision. ZD and HT: reviewing the literature and analyzing. DK: making figures and tables. FX: rechecking the manuscript and putting forward suggestions for amendment. NL: revising language and content, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key Laboratory of Sports Medicine of Sichuan Province, Institute of Sports Medicine and Health, Chengdu Sport University (No. 2021-A030) and the National Key Research and Development Program of China (No. 2018YFF0300904).

REFERENCES

1. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Internet Res.* (2019) 21:e13216. doi: 10.2196/13216
2. Akbari A, Haghverd F, Behbahani S. Robotic home-based rehabilitation systems design: from a literature review to a conceptual framework for community-based remote therapy during COVID-19 pandemic. *Front Robot AI.* (2021) 8:612331. doi: 10.3389/frobt.2021.612331
3. Kaelin VC, Valizadeh M, Salgado Z, Parde N, Khetani MA. Artificial intelligence in rehabilitation targeting the participation of children and youth with disabilities: scoping review. *J Med Internet Res.* (2021) 23:e25745. doi: 10.2196/25745
4. Li G, Zheng Y, Wu X, Hu Y, Fang P, Xiong J, Xia Z, Wang C. Research progress and trend of medical rehabilitation robot. *Proc Chin Acad Sci USA*. (2015) 30:793–802. doi: 10.16418/j.issn.1000-3045.2015.06.013
5. Yu H. Rehabilitation robots: ten visions for the future. *Chin J Rehabil Med*. (2020) 35:900–2. doi: 10.3969/j.issn.1001-1242.2020.08.002
6. Wang Q, Wei Y, Liu L. Progress in research and application of rehabilitation robots. *Pack Eng*. (2018) 39:83–9. doi: 10.19554/j.cnki.1001-3563.2018.18.018
7. Li H, Zhang T, Feng Y, Lin H, Bai D. Application of exoskeleton-based lower limb rehabilitation robot in stroke rehabilitation. *Chin Rehabil Theory Pract*. (2017) 023(007):788-791. doi: 10.3969/j.issn.1006-9771.2017.07.010
8. Butchart J, Harrison R, Ritchie J, Martí F, McCarthy C, Knight S, et al. Child and parent perceptions of acceptability and therapeutic value of a socially assistive robot used during pediatric rehabilitation. *Disabil Rehabil.* (2021) 43:163–170. doi: 10.1080/09638288.2019.1617357

9. Fang J, Schuwey A, Stocker N, Pedrini B, Sampaio A, Hunt KJ. Preliminary development and technical evaluation of a belt-actuated robotic rehabilitation platform. *Technol Health Care.* (2021) 29:595–607. doi: 10.3233/THC-202392
10. Nam C, Rong W, Li W, Cheung C, Ngai W, Cheung T, et al. An exoneuromusculoskeleton for self-help upper limb rehabilitation after stroke. *Soft Robot.* (2020). doi: 10.1089/soro.2020.0090
11. Pérez-San Lázaro R, Salgado I, Chairez I. Adaptive sliding-mode controller of a lower limb mobile exoskeleton for active rehabilitation. *ISA Trans.* (2021) 109:218–228. doi: 10.1016/j.isatra.2020.10.008
12. Rodgers H, Bosomworth H, Krebs HI, van Wijck F, Howel D, Wilson N, et al. Robot assisted training for the upper limb after stroke (RATULS): a multicentre randomised controlled trial. *Lancet.* (2019) 394:51–62. doi: 10.1016/S0140-6736(19)31055-4
13. Bernhardt J, Mehrholz J. Robotic-assisted training after stroke: RATULS advances science. *Lancet.* (2019) 394:6–8. doi: 10.1016/S0140-6736(19)31156-0
14. Li J. Human-machine integration, nature and man integration—thoughts on the application and development of rehabilitation robots. *Chin J Rehabil Med.* (2020) 35:897–9. doi: 10.3969/j.issn.1001-1242.2020.08.001
15. Mehrholz J, Hädrich A, Platz T, Kugler J, Pohl M. Electromechanical and robot-assisted arm training for improving generic activities of daily living, arm function, and arm muscle strength after stroke. *Cochrane Database Syst Rev.* (2012) 6:CD006876. doi: 10.1002/14651858.CD006876.pub3
16. Nam KY, Kim HJ, Kwon BS, Park JW, Lee HJ, Yoo A. Robot-assisted gait training (Lokomat) improves walking function and activity in people with spinal cord injury: a systematic review. *J Neuroeng Rehabil.* (2017) 14:24. doi: 10.1186/s12984-017-0232-3
17. Lefmann S, Russo R, Hillier S. The effectiveness of robotic-assisted gait training for paediatric gait disorders: systematic review. *J Neuroeng Rehabil.* (2017) 14:1. doi: 10.1186/s12984-016-0214-x
18. Krebs HI, Volpe BT. Rehabilitation robotics. *Handbook Clin Neurol.* (2013) 110:283–94. doi: 10.1016/B978-0-444-52901-5.00023-X
19. Kim WS, Cho S, Ku J, Kim Y, Lee K, Hwang HJ, et al. Clinical application of virtual reality for upper limb motor rehabilitation in stroke: review of technologies and clinical evidence. *J Clin Med.* (2020) 9:3369. doi: 10.3390/jcm9103369
20. Keeling AB, Piitz M, Semrau JA, Hill MD, Scott SH, Dukelow SP. Robot enhanced stroke therapy optimizes rehabilitation (RESTORE): a pilot study. *J Neuroeng Rehabil.* (2021) 18:10. doi: 10.1186/s12984-021-00804-8
21. Chien WT, Chong YY, Tse MK, Chien CW, Cheng HY. Robot-assisted therapy for upper-limb rehabilitation in subacute stroke patients: a systematic review and meta-analysis. *Brain Behav.* (2020) 10:e01742. doi: 10.1002/brb3.1742
22. Ranzani R, Lambercy O, Metzger JC, Califfi A, Regazzi S, Dinacci D, et al. Neurocognitive robot-assisted rehabilitation of hand function: a randomized control trial on motor recovery in subacute stroke. *J Neuroeng Rehabil.* (2020) 17:115. doi: 10.1186/s12984-020-00746-7
23. Synnestvedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. In: *AMIA Annual Symposium Proceedings AMIA Symposium*. (2005), p. 724–8.
24. Chen C, Chen Y. Searching for clinical evidence in CiteSpace. In: *AMIA Annual Symposium Proceedings AMIA Symposium*. (2005), p. 121–5.
25. Chaomei C, Zhigang H, Shengbo L, Hung T. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. *Expert Opin Biol Ther.* (2013) 2012:593–608. doi: 10.1517/14712598.2012.674507
26. Sun W, Huang P, Song H, Feng D. Bibliometric analysis of acute pancreatitis in Web of Science database based on CiteSpace software. *Medicine.* (2020) 99:e23208. doi: 10.1097/MD.0000000000023208
27. Lu C, Li X, Yang K. Trends in shared decision-making studies from 2009 to 2018: a bibliometric analysis. *Front. Public Health.* (2019) 7:384. doi: 10.3389/fpubh.2019.00384
28. van Hedel HJA, Severini G, Scarton A, O'Brien A, Reed T, Gaebler-Spira D, et al. Advanced Robotic Therapy Integrated Centers (ARTIC): an international collaboration facilitating the application of rehabilitation technologies. *J Neuroeng Rehabil.* (2018) 15:30. doi: 10.1186/s12984-018-0378-7
29. Liang C, Luo A, Zhong Z. Knowledge mapping of medication literacy study: A visualized analysis using CiteSpace. *SAGE Open Med.* (2018) 6:2050312118800199. doi: 10.1177/2050312118800199
30. Ping X. Study of international anticancer research trends via co-word and document co-citation visualization analysis. *Scientometrics.* (2015) 105:611–22. doi: 10.1007/s11192-015-1689-0
31. Xia D, Yao R, Wang S, Chen G, Wang Y. Mapping trends and hotspots regarding clinical research on COVID-19: a bibliometric analysis of global research. *Front. Public Health.* (2021) 9:713487. doi: 10.3389/fpubh.2021.713487
32. Langhorne PP, Bernhardt PJ, Kwakkel G. Stroke care 2: stroke rehabilitation. *Lancet.* (2011) 377:1693–702. doi: 10.1016/S0140-6736(11)60325-5
33. Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, et al. Robot-assisted therapy for long-term upper-limb impairment after stroke. *N Engl J Med.* (2010) 362:1772–83. doi: 10.1056/NEJMoa0911341
34. Polygerinos P, Wang Z, Galloway KC, Wood RJ, Walsh CJ. Soft robotic glove for combined assistance and at-home rehabilitation. *Robot Autonom Syst.* (2014) 73(C):135–43. doi: 10.1016/j.robot.2014.08.014
35. Esquenazi A, Talaty M, Packel A, Saulino M. The ReWalk powered exoskeleton to restore ambulatory function to individuals with thoracic-level motor-complete spinal cord injury. *Am J Phys Med Rehabil.* (2012) 91:911–21. doi: 10.1097/PHM.0b013e318269d9a3
36. Krakauer JW, Carmichael ST, Corbett D, Wittenberg GF. Getting neurorehabilitation right: what can be learned from animal models? *Neurorehabil Neural Repair.* (2012) 26:923. doi: 10.1177/1545968312440745
37. Park J, Lee Y, Hong J, Lee Y, Ha M, Jung Y, et al. Tactile-direction-sensitive and stretchable electronic skins based on human-skin-inspired interlocked microstructures. *ACS Nano.* (2014) 8:12020–9. doi: 10.1021/nn505953t
38. Kiguchi K, Hayashi Y. An EMG-based control for an upper-limb power-assist exoskeleton robot. *IEEE Trans Syst Man Cybern Part B.* (2012) 42:1064–71. doi: 10.1109/TSMCB.2012.2185843
39. Klamroth-Marganska V, Blanco J, Campen K, Curt A, Dietz V, Ettlin T, et al. Three-dimensional, task-specific robot therapy of the arm after stroke: a multicentre, parallel-group randomised trial. *Lancet Neurol.* (2014) 13:159–66. doi: 10.1016/S1474-4422(13)70305-3
40. Park YL, Chen BR, Pérez-Arancibia NO, Young D, Stirling L, Wood RJ, et al. Design and control of a bio-inspired soft wearable robotic device for ankle-foot rehabilitation. *Bioinspir Biomimet.* (2014) 9:016007. doi: 10.1088/1748-3182/9/1/016007
41. Heo P, Gu GM, Lee SJ, Rhee K, Kim J. Current hand exoskeleton technologies for rehabilitation and assistive engineering. *Int J Precis Eng Manuf.* (2012) 13:807–824. doi: 10.1007/s12541-012-0107-2
42. Maciejasz P, Eschweiler J, Gerlach-Hahn K, Jansen-Troy A, Leonhardt S. A survey on robotic devices for upper limb rehabilitation. *J Neuroeng Rehabil.* (2014) 11:1–29. doi: 10.1186/1743-0003-11-3
43. Prior SD, Warner PR. A review of world rehabilitation robotics research. In: *IEEE Colloquium on High-tech Help for the Handicapped* (2002).
44. Musk E. An integrated brain-machine interface platform with thousands of channels. *J Med Internet Res.* (2019) 21:e16194. doi: 10.2196/16194
45. Singh HR, Mobin A, Kumar S, Chauhan S, Agrawal SS. Design and development of voice/joystick operated microcontroller based intelligent motorised wheelchair. In: *Tencon 99 IEEE Region 10 Conference* (1999).
46. Price DDS. Little science big science. *Paperback.* (1963)
47. Ahmad P, Asif JA, Alam MK, Slots J. A bibliometric analysis of Periodontology 2000. *Periodontology.* (2000) 82:286–297. doi: 10.1111/prd.12328
48. He J, He L, Geng B, Xia Y. Bibliometric analysis of the top-cited articles on unicompartmental knee arthroplasty. *J Arthroplasty.* (2021) 36:1810–8.e3. doi: 10.1016/j.arth.2020.11.038
49. Mehrholz J, Pollock A, Pohl M, Kugler J, Elsner B. Systematic review with network meta-analysis of randomized controlled trials of robotic-assisted arm training for improving activities of daily living and upper limb function after stroke. *J Neuroeng Rehabil.* (2020) 17:83. doi: 10.1186/s12984-020-00715-0
50. Schicketmueller A, Lamprecht J, Hofmann M, Sailer M, Rose G. Gait event detection for stroke patients during robot-assisted gait training. *Sensors.* (2020) 20:3399. doi: 10.3390/s20123399
51. Blaya JA, Herr H. Adaptive control of a variable-impedance ankle-foot orthosis to assist drop-foot gait. *IEEE Trans Neural Syst Rehabil Eng.* (2004) 12:24–31. doi: 10.1109/TNSRE.2003.823266
52. Thippa RG, Praveen K, Lakshmanna K, Kaluri R, Baker T. Analysis of dimensionality reduction techniques on big data. *IEEE Access.* (2020) 8:54776–88. doi: 10.1109/ACCESS.2020.2980942

53. Sarhan MH, Nasseri MA, Zapp D, Maier M, Lohmann CP, Navab N, et al. Machine learning techniques for ophthalmic data processing: a review. *IEEE J Biomed Health Inform.* (2020) 24:3338–50. doi: 10.1109/JBHI.2020.3012134
54. Liu Y, Zhang R, Nie F, Li X, Ding C. Supervised dimensionality reduction methods via recursive regression. *IEEE Trans Neural Netw Learning Syst.* (2020) 31:3269–79. doi: 10.1109/TNNLS.2019.2940088
55. Seo K, Chung B, Panchaseelan HP, Kim T, Park H, Oh B, et al. Forecasting the walking assistance rehabilitation level of stroke patients using artificial intelligence. *Diagnostics.* (2021) 11:1096. doi: 10.3390/diagnostics11061096
56. Scrutinio D, Ricciardi C, Donisi L, Losavio E, Battista P, Guida P, et al. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Sci Rep.* (2020) 10:20127. doi: 10.1038/s41598-020-77243-3
57. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* (2018) 374:20150202. doi: 10.1038/nbt.4314
58. Sun X, Liu Y, An L. Ensemble dimensionality reduction and feature gene extraction for single-cell RNA-seq data. *Nat Commun.* (2020) 11:5853. doi: 10.1038/s41467-020-19465-7
59. Page RJ. *A New Implant for Blind People Jacks Directly into the Brain*. MIT Technology Review. (2020). Available online at: https://www.technologyreview.com/2020/02/06/844908/a-new-implant-for-blind-people-jacks-directly-into-the-brain/ (accessed February 6, 2020).
60. Zhang X, Chan FK, Parthasarathy T, Gazzola M. Modeling and simulation of complex dynamic musculoskeletal architectures. *Nat Commun.* (2019) 10:4825. doi: 10.1038/s41467-019-12759-5
61. Xue X, Tu H, Deng Z, Zhou L, Li N, Wang X. Effects of brain-computer interface training on upper limb function recovery in stroke patients: a protocol for systematic review and meta-analysis. *Medicine.* (2021) 100:e26254. doi: 10.1097/MD.0000000000026254
62. Abbruzzese G, Marchese R, Avanzino L, Pelosin E. Rehabilitation for Parkinson's disease: current outlook and future challenges. *Parkinsonism Relat Disord.* (2016) 22 Suppl 1:S60–4. doi: 10.1016/j.parkreldis.2015.09.005
63. Khor KX, Chin PJH, Yeong CF, Su ELM, Narayanan ALT, Abdul Rahman H, et al. Portable and reconfigurable wrist robot improves hand function for post-stroke subjects. *IEEE Trans Neural Syst Rehabil Eng.* (2017) 25:1864–73. doi: 10.1109/TNSRE.2017.2692520
64. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. *JMIR mHealth uHealth.* (2020) 8:e18907. doi: 10.2196/18907
65. Treviño LR, Roberge P, Auer ME, Morales A, Torres-Reveron A. Predictors of functional outcome in a cohort of hispanic patients using exoskeleton rehabilitation for cerebrovascular accidents and traumatic brain injury. *Front Neurorobot.* (2021) 15:682156. doi: 10.3389/fnbot.2021.682156
66. Hussain S, Jamwal PK, Vliet PV, Brown NAT. Robot assisted ankle neuro-rehabilitation: state of the art and future challenges. *Exp Rev Neurotherap.* (2021) 21:111–21. doi: 10.1080/14737175.2021.1847646
67. Klamroth-Marganska V. Stroke rehabilitation: therapy robots and assistive devices. *Adv Exp Med Biol.* (2018) 1065:579–87. doi: 10.1007/978-3-319-77932-4_35

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xue, Yang, Deng, Tu, Kong, Li and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fpubh.2021.824898

A Hybrid Framework for Intrusion Detection in Healthcare Systems Using Deep Learning

M. Akshay Kumaar[1], Duraimurugan Samiayya[2], P. M. Durai Raj Vincent[3], Kathiravan Srinivasan[4], Chuan-Yu Chang[5,6] and Harish Ganesh[2]*

[1] BrainSightAI, Bangalore, India, [2] Department of Information Technology, St. Joseph's College of Engineering, Chennai, India, [3] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [4] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India, [5] Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu, Taiwan, [6] Service Systems Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Abdul Rehman Javed,
Air University, Pakistan
Suresh Manic Kesavan,
National University of Science and Technology, Oman
Aswathy Su,
Jyothi Engineering College, India

***Correspondence:**
Chuan-Yu Chang
chuanyu@yuntech.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 29 November 2021
Accepted: 16 December 2021
Published: 12 January 2022

Citation:
Akshay Kumaar M, Samiayya D, Vincent PMDR, Srinivasan K, Chang C-Y and Ganesh H (2022) A Hybrid Framework for Intrusion Detection in Healthcare Systems Using Deep Learning. Front. Public Health 9:824898. doi: 10.3389/fpubh.2021.824898

The unbounded increase in network traffic and user data has made it difficult for network intrusion detection systems to be abreast and perform well. Intrusion Systems are crucial in e-healthcare since the patients' medical records should be kept highly secure, confidential, and accurate. Any change in the actual patient data can lead to errors in the diagnosis and treatment. Most of the existing artificial intelligence-based systems are trained on outdated intrusion detection repositories, which can produce more false positives and require retraining the algorithm from scratch to support new attacks. These processes also make it challenging to secure patient records in medical systems as the intrusion detection mechanisms can become frequently obsolete. This paper proposes a hybrid framework using Deep Learning named "ImmuneNet" to recognize the latest intrusion attacks and defend healthcare data. The proposed framework uses multiple feature engineering processes, oversampling methods to improve class balance, and hyper-parameter optimization techniques to achieve high accuracy and performance. The architecture contains <1 million parameters, making it lightweight, fast, and IoT-friendly, suitable for deploying the IDS on medical devices and healthcare systems. The performance of ImmuneNet was benchmarked against several other machine learning algorithms on the Canadian Institute for Cybersecurity's Intrusion Detection System 2017, 2018, and Bell DNS 2021 datasets which contain extensive real-time and latest cyber attack data. Out of all the experiments, ImmuneNet performed the best on the CIC Bell DNS 2021 dataset with about 99.19% accuracy, 99.22% precision, 99.19% recall, and 99.2% ROC-AUC scores, which are comparatively better and up-to-date than other existing approaches in classifying between requests that are normal, intrusion, and other cyber attacks.

Keywords: healthcare, network security, neural networks, artificial intelligence, big data, deep learning

INTRODUCTION

Network intrusion recognition is challenging since the attacks evolve daily because of new technologies, frameworks, and software. In 2020, the number of cyber-attacks increased by 17%, in that 77% were targeted attacks, with attackers' main targets being personal data and credentials. Attacks on organizations aimed mainly at stealing private user data. These metrics show a vital backdrop in modern-day cyber-attack detection and prevention. Moreover, the healthcare industry is increasing, and most hospitals are integrating e-healthcare systems to meet the patients' needs as soon as possible. Hospitals must maintain the Electronic Health Records (EHR) and Patient Records or Personal Health Records (1) since these details contain a patient's medical data required to infer a diagnosis and treatment. The vast development in the Internet of Things (IoT) has led to a boom in smart medical devices and systems. These edge devices can contain patient records, which must be kept secure and accurate at all times. Any change or corruption in these details can lead to wrong diagnosis and treatment, causing the fatality of the patient. Therefore, to ensure cyber-safety in healthcare systems, there is a need for up-to-date and advanced Hybrid Intrusion Detection Systems.

The common goal of any Intrusion Detection System is to recognize, flag, and log/block intrusion attacks by identifying any malicious network activity (1, 2). Most of the existing real-time software for IDS uses a rule-based approach like signature-based detection, stateful protocol analysis, and statistical packet analysis. Primarily the IDS classifies a request into benign and malicious, benign being regular requests, and malicious being anomalous or intrusion requests. The IDS are also specifically designed to identify a specific set of attacks like DDoS.

However, Artificial Intelligence-based systems come with specific backlogs; the rate of false positives can be overwhelmingly frequent compared to actual threats. The algorithms may also be biased toward particular attacks based on class imbalance and features present in the dataset used for training the algorithm. Inaccurate and biased intrusion detection systems don't stand a chance in protecting patient records as they may falsely flag an attack as benign or a benign request as malicious. The issue with the existing intrusion detection systems is that they are trained on outdated or old datasets like KDD-Cup'99, making the system vulnerable to the latest attacks and leading to leakage and illegal modification of EHRs. Traditional Machine Learning algorithms such as K-Nearest Neighbors (3) and Logistic Regression usually need repeated fine-tuning and updating parameters to adapt the algorithm to new types of attacks that may require frequent retraining, which is computationally expensive and time-consuming. Ensemble learning (4) and deep learning-based (5) approaches have proven better because of their adaptability and ease of fine-tuning or pruning on new data. Faster and efficient algorithmic techniques are still vital for such critical applications as time complexity is a significant overhead in medical and IoT-based systems. A state-of-the-art, present-day attack-oriented, faster, and efficient methodology is required for Intrusion Detection Systems to safeguard patient records in e-healthcare.

This paper proposes a new hybrid framework for intrusion detection using deep learning for healthcare systems named "ImmuneNet." We have benchmarked its performance against various machine learning algorithms on the Canadian Institute for Cybersecurity's IDS 2017 (6), IDS 2018 (7), Bell DNS 2021 (8) datasets. These are realistic datasets containing more than 17 types of the latest cyberattacks, which would be beneficial in implementing real-time deep learning-based intrusion detection systems and achieving high accuracy. A data-centric approach was used to promote class balance (9) and extensive feature selection processes like recursive feature elimination (10, 11), which can help better perform models with fewer parameters. Over five types of machine learning algorithms were compared: Logistic Regression, Decision Trees, Random Forests, Extreme Gradient Boosting Trees (12), and a custom neural network architecture named "ImmuneNet" as seen in **Figure 1**. ImmuneNet obtained the highest accuracy of almost 99.19% and other vital metrics from all the experiments. The neural network architecture contains around 830,000 parameters, making the fine-tuning process faster and deployment-friendly for medical devices and systems.

Our experiments were based on several data-centric and algorithmic approaches to create a lightweight, fast, high-performance classifier without compromising accuracy. We have divided our paper into the following sections based on different schemes:

1. The current state-of-the-art methodologies and existing research for implementing intrusion detection systems in healthcare and general.
2. Comparison between KDD Cup99, CIC-IDS variants, and CIC Bell DNS 2021 datasets to know in detail about the attack types, size, number of samples, and state of relevance.
3. Extensive feature selection, oversampling, and hyper-parameter optimization techniques involved in training, validating, and testing the algorithms on the three datasets.
4. Comparison of multiple machine learning algorithms, namely logistic regression, ensemble methods such as decision trees, extreme gradient boosting trees, and a custom neural network architecture named "ImmuneNet" using various metrics like accuracy, loss, precision, recall, F1, and loss.
5. A detailed discussion on the proposed system's performance, advantages, and limitations.

RELATED WORKS

Our literature survey primarily focused on analyzing the data preprocessing and algorithmic approaches to derive new novel solutions for developing a sound intrusion detection system. There are tons of cyber vulnerabilities to cope within the medical sector as these attacks threaten patients' healthcare and put them in jeopardy. Tervoot et al. (13) performed a scoping review by categorizing and analyzing these attacks and found 18 solutions, each fitting at least one of the categories of intrusion detection and prevention, communication tunneling, or hardware protections. Thamilarasu et al. (14) proposed an intrusion detection system for the internet of medical things

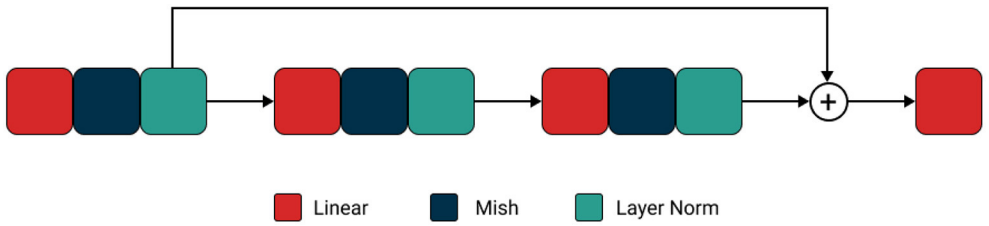


FIGURE 1 | Neural network of ImmuneNet architecture.

based on Wireless body area networks (WBAN). They simulated a hospital network topology and performed detailed experiments for various subsets of the Internet of Medical Things, including wireless body area networks and other connected medical devices. The system performs best at 99.6 and 98.2% accuracy for network and device-level intrusion detection, respectively.

Subas et al. (4) proposed an IDS with multiple Machine Learning techniques with some complex attack data. They concluded that the Bagging ensemble algorithm and Random Forest perform better with 97.67% accuracy than other classifier models. Hence, a Smart E-Healthcare System could be built on top of this system to ensure better cyber security. The Internet of Medical Things (IMoT) has its drawbacks as technical advancements progress since any security and privacy issues exist. Therefore, Priya et al. (5) developed a deep neural network (DNN) for constructing an effective IDS to classify and predict unforeseen cyberattacks. They also carried out practical feature engineering in hyperparameter selection. The proposed DNN model performs better than the existing machine learning approaches, with an increase in accuracy by 15% with an overall accuracy of 99.7 % on NSL-KDD. Šabić et al. (15) proposed an anomaly detection system that focuses on the system's ability to detect anomalies in heart rate data. They used five algorithms to fit the data, two of them were unsupervised, and the rest were supervised to detect the anomalies. They concluded that random Forest and ensemble methods had above 99% accuracy to model such systems effectively.

Hady et al. (16) proposed and built a real-time Enhanced Healthcare Monitoring System (EHMS) that monitors the patients' biometrics and collects network flow metrics. Then this system applies different machine learning methods for training and testing the dataset against these attacks. Hence by collecting their data dynamically, they developed a robust intrusion detection system. by analyzing in 10-fold accuracy score comparison, they noted that the SVM algorithm performed better than the rest with an accuracy of about 92.44%. Nguyen et al. (17) used data fusion as an integral part of multidisciplinary research to design an intrusion detection system. Their proposed model involves a decision-based fusion model with different processes such as initialization, preprocessing, Feature Selection, and multimodal classification to detect intrusions effectively. The proposed model offers maximum accuracy of 99.21%, precision of 98.93%, and a detection rate of 99.59%. Iwendi et al. (18) proposed an intrusion detection system in the Security of Things (SoT) paradigm for smart healthcare and will continue to impact medical infrastructures. Their study used the NSL-KDD dataset on RF, Naive Bayes (NB), and logistic regression classifiers for machine learning. To optimize the functionality of their approach, they used a weighted genetic algorithm, and Random Forest was coupled to generate achieved a high detection rate and neglect false positives and true negatives. The combination of their genetic algorithm and RF models achieved an accuracy rate of 98.81%.

Yeng et al. (3) conducted a systematic review of various machine learning algorithms and data resources to assess them across multiple criteria. They have also evaluated and analyzed the design considerations of the methods toward mitigating false positives. Their comprehensive study determines the associated challenges in using the algorithms and how they can be overcome. Their study suggests that most of the existing literature for intrusion systems in healthcare uses K-Nearest Neighbors. Mahdavifar et al. (8) have proposed a method to classify benign, phishing, spam, and malware-based domains using DNS traffic analysis with the K-Nearest Neighbors algorithm. They have used the CIC Bell DNS 2021 dataset, which is balanced to 60/50 and 97/3%, achieving an F1 score of around 94.8 and 99.4%, respectively. Seth et al. (19) built a unique framework based on creating an ensemble by ranking the detection ability of different base classifiers to identify various types of attacks. They have experimented with numerous algorithms and achieved 96.65% accuracy with neural networks on the CIC-IDS 2018 dataset and have also given a deep analysis of the metrics on classwise performance. Benedetto Serinellia et al. (20) analyses open-source intrusion detection datasets by validating them to detect well-known zero-day attacks. They have built their predictors based on KDD99, NSL-KDD, and CIC-IDS-2018 datasets, giving a comparative case study among them in their research work. They achieved over 99.97% accuracy on the NSL-KDD datasets. Thilagam et al. (21), on the other hand, proposed an Intrusion detection system on a more complex basis using RCNN (Recurrent Convolutional Neural Networks) using the culmination of CNN and LSTM (Long Short Term Memory) on the KDD Cup 99 and CIC-IDS 2018 datasets. They achieved over 94% accuracy on both datasets with the same model.

Class Imbalance is another critical attribute to be considered for building a novel and efficient architecture. Zuech et al. (22) analyzed web attacks using random undersampling ratios under various ensemble learning algorithms and discussed

TABLE 1 | Comparison of KDDCup99, CIC IDS variants, and CIC Bell DNS 2021 datasets.

	KDD Cup 99	CIC IDS 2017	CIC IDS 2018	CIC Bell DNS 2021
Year	1999	2017	2018	2021
Provider	University of California, Irvine	Canadian Institute for Cybersecurity, University of New Brunswick	Canadian Institute for Cybersecurity, University of New Brunswick	Canadian Institute for Cybersecurity, University of New Brunswick + Bell, Canada + Cyber Threat Intelligence, Canada
Size (GB)	0.75	1.06	5.81	0.287
Number of samples	494,020	2,830,743	16,232,943	413,011
Classes	DoS, U2R, R2L, Probe + sub-attacks	DDoS, DoS slowloris, BotNet, Port Scan, SQL Injection + 7	DDoS, DoS slowloris, BotNet, Port Scan, SQL Injection + 10	Benign, Spam, Phishing, Malware

TABLE 2 | Features in CIC IDS 2018 with corresponding correlation values and *p*-value based significance.

S. No	Feature name	Correlation value	*p*-value
1	active_mean	0.9086	>0.05
2	Protocol	0.7950	<0.05
3	subflow_fwd_pkts	0.9992	>0.05
4	flow_pkts_s	0.2177	<0.05
5	bwd_pkts_s	0.2349	<0.05
6	pkt_size_avg	0.9410	>0.05

class balance's significant importance. They observed that undersampling at different ratios could have a drastic effect on the model's performance and achieved an accuracy of about 94.01% accuracy on CIC IDS 2018 dataset using RCNN. Yu et al. (23) had proposed a packet byte-based CNN, called PBCNN, which focuses on the statistical features of network traffic, giving another insight on the importance of network traffic and its hierarchical nature. PBCNN obtained 99.99 % accuracy on the CIC IDS 2017 and CIC IDS 2018 datasets. Gopalan et al. (9) carried out a balancing approach and surveyed the CIC-IDS-2018 dataset. They also researched the impact of bias and class imbalances in the CIC-IDS-2018 dataset. As observed, signature-based intrusion detection approaches prove to be the weaker counterpart of various methods.

Most of the research on Intrusion detection mechanisms was based on the ensemble learning approach. Fitni et al. (24) made comparisons with seven single classifiers to identify the most appropriate basic classifiers for ensemble learning; they compared the accuracy metrics of tested architectures and made a cumulative study on it. They also used effective feature selection techniques like filter methods and Spearman rank's correlation and used only 23 features to achieve 98.8% accuracy on the CIC IDS 2018 dataset. We used the MISH activation function in our custom neural network architecture. Misra (25), the author of MISH. Discussed in detail its scope and found that. Mish outperformed other activation functions like Leaky ReLU on YOLOv4 with a CSP-DarkNet-53 backbone on average precision. Hua et al. (26) proposed using the LightGBM ensemble method, which proved to be quite a popular and efficient ensemble method for Intrusion Detection. By achieving 98.37% accuracy, the study also compared their approach with other ensemble methods and traditional machine learning algorithms to prove its efficiency.

Catillo et al. (27) presented an autoencoder-based bi-level anomaly detector, which used a specific Artificial Neural Network to reconstruct their vectorized input on the CIC IDS 2017 and 2018 datasets with a maximum detection rate of 99.2 %. It seems to be an overwhelming algorithmic approach to a relatively simple problem with minimal feature space. Khan and Kim (28) have proposed a Convolutional auto-encoder (Conv-AE) based Intrusion Detection System using a heterogeneous dataset. The Conv-AE algorithm proves to efficiently combine the advantages of traditional state-of-art approaches and identifies malicious attacks in the case of both anomaly and misused-based ID systems. They have also mentioned the distributed and big-data-friendly processing pipelines for training their algorithms and achieved 98.2 % accuracy. Meryem and Ouahidi (29) overcame the limitations of anomaly and misuse detection by using a hybrid approach. They have employed the K-Nearest Neighbors on the KDD-Cup99 dataset. They had achieved more than 98.77% and an Area Under the Curve (AUC) score of 0.58, which is sub-optimal. Thakkar and Lohiya (30) briefly discuss CIC-IDS-2017 and CIC-IDS-2018 datasets by analyzing their features and iterating them across many Machine Learning and Deep Learning models. They had suggested that the CIC IDS datasets were significantly better than NSL KDD because of the type of attacks, data capturing techniques, and attack infrastructure.

According to Gao et al. (31), an adaptive and ensemble system to the same anomaly-based approach proved better. Using the KDD Cup99 dataset, they used multiple algorithms such as KNN, DNN, random forest, and to architect an ensemble adaptive voting algorithm. They compared the algorithms and found that the Adaboost model achieved 99.99% accuracy. Reviewing and understanding the dataset is vital for building innovative and novel solutions. Chawla et al. (32) in their paper, had attempted to build an efficient intrusion detection system that was anomaly-based, using Recurrent Neural Networks like Gated Recurrent Units rather than the vanilla LSTM networks. However, convolutions and recurrent methods can be computationally intensive and time-consuming compared to other methodologies applied for non-sequential tabular data.

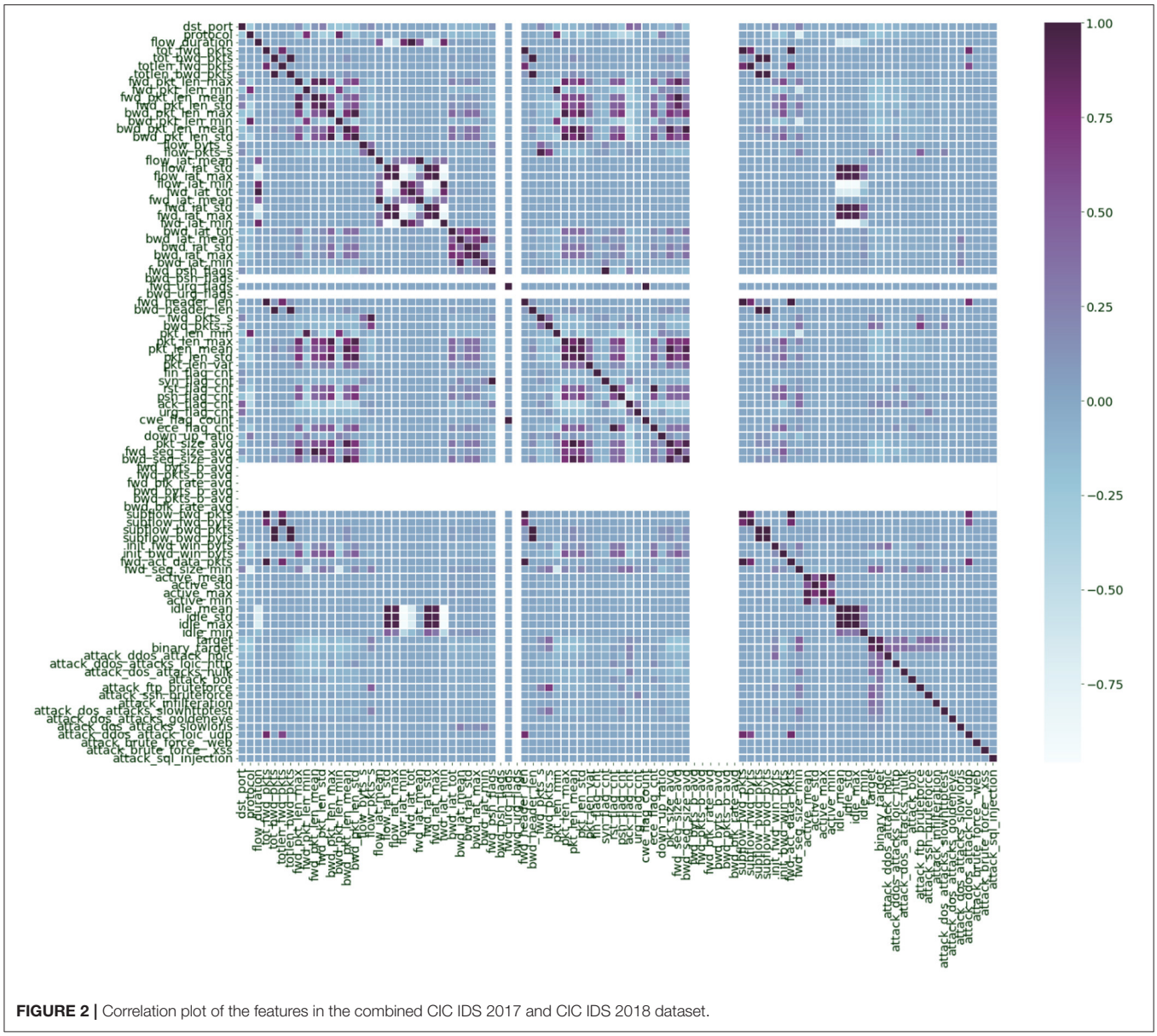


FIGURE 2 | Correlation plot of the features in the combined CIC IDS 2017 and CIC IDS 2018 dataset.

They have obtained an accuracy of only around 81% on the ADFA Intrusion Detection dataset. Anomaly-based detection has its disadvantages, supporting the opinion of Feras et al. (33) proposed a hybrid model that leverages the functionalities of SVMs and decision trees. Using the CIC IDS 2018 and KDD Cup 99 datasets, they have accuracy of 97.881 and 99.982%, respectively. Wankhede and Kshirsagar (34)–"DoS Attack Detection Using Machine Learning and Neural Network" uses the dataset CIC IDS 2017 dataset and uses Random Forest and Multi-Layer Perceptron approach. The Random Forest algorithm provides more optimized results than Multi-Layer Perceptron from their experiments, obtained around 99% accuracy on the CIC-IDS 2017 datasets.

Mubashar et al. (1) created a framework for an IoT-based medical data archival system by combining a novel block chain-based technique for medical data encryption with an optimization algorithm. Loshchilov and Hutter (35) proposed Stochastic Gradient Descent with Warm Restarts, from which

TABLE 3 | Example features in CIC Bell DNS 2021 dataset.

S. No	Country	ASN	Entropy	Name_Server _Count	Domain_Age
1	NL	20,415.0	3.1547	0.0	1,735
2	PL	198,414.0	2.6250	4.0	5,117
3	US	46,606.0	3.0079	4.0	510
4	AU	45,638.0	3.4058	17.0	307
5	FR	12,876.0	2.2893	25.0	603

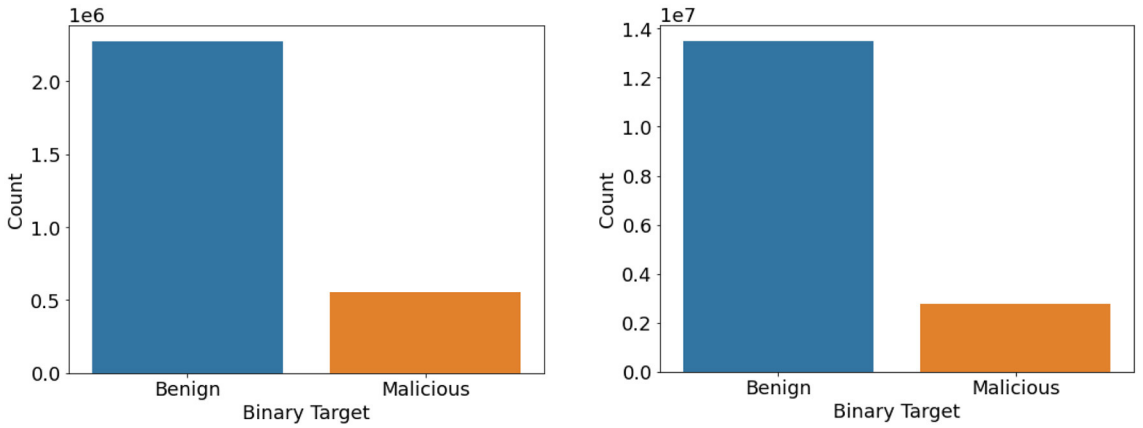


FIGURE 3 | Binary class count plot of CIC IDS 2017 (left) and 2018 (right) datasets.

we have used the Cosine annealing learning rate scheduling technique as an optimization strategy in our experiments. Glorot and Bengio (36) had proposed a new weights initialization technique for neural networks that leads to better convergence and performance of the model. Sharma and Yadav (10) in their paper have proposed recursive feature elimination technique for feature selection and have achieved an accuracy of about 99% accuracy for DOS attacks and other attacks using Decision Trees on KDD Cup 99 dataset. Correlation based feature selection was proposed by Thaseen et al. (37) in their paper as a feature selection method, and obtained an overall accuracy of 97.9 % with an artificial neural network on the NSL-KDD dataset (38–42).

From our observation, most of the research used the KDD-Cup99, and the approaches done with CIC-IDS datasets support the latest intrusion attacks. They are performing better while also supporting better classes of attacks. The features in the CIC-IDS 2017 and 2018 datasets are quite the same. They can also be used together with multiple feature selection techniques and sampling techniques to produce better results for the classifiers. The Intrusion Detection mechanism must be trained on the latest cyber-attack repositories to be up-to-date hence protecting the patient records without any compromise. It is also essential that the algorithm is not biased toward benign or malicious classes, which may cause high false positive rates (43–50). Hence, a balanced dataset should be used for training the IDS mechanism. Our paper has proposed the methodologies followed for feature selection, class balancing, data preparation, and preprocessing under multiple sub-sections in section Methodology. The results, comparison for the algorithms, and discussion on the pros and cons of the proposed system are discussed under sections Results and Discussion, respectively.

METHODOLOGY

Our approach deals with building a hybrid intrusion detection mechanism that can support the latest cyberattacks without following any traditional rule-based methods. The IDS should protect electronic Health Records (EHR) from any type of attack. An algorithm that can be pruned or fine-tuned to adapt to new attacks is crucial since recent attacks are evolving daily. We experimented with a new dataset by the Canadian Institute for Cybersecurity that supports over 15 types of latest intrusion attacks and compared the same with the widely available datasets for intrusion detection. A statistical feature selection process has been done on the dataset to extract the most valuable features that can boost the performance of the classifiers. We also dealt with the class imbalance of the dataset to not make our model biased toward particular positives or negatives. We compared five machine learning algorithms: Logistic Regression, Decision Trees, random forests, XGB (12), and Artificial Neural Network. We have created a new Artificial Neural Network architecture that surpassed performance compared to the other algorithms obtaining ∼99.2 % accuracy. We conducted our experiments on a Linux machine running Ubuntu 21.04 OS with 32 GB RAM and 8 GB NVidia GTX 1080 GPU to use distributed training of our algorithms to obtain results faster. We used Python and popular data science libraries like NumPy, Pandas, Scikit-Learn, and TensorFlow to implement the data processing and algorithms. We have discussed all the aspects of our methodology in detail in the following subsections.

Dataset Comparison

Datasets used for Intrusion Detection and Coping purposes evolve a lot with time since there is a perennial and dynamic change in cybersecurity threats and new attacks rise around the horizon day in and out. The most traditional dataset used for this purpose is the KDD Cup 99 dataset, which is proven to have a lot of anomalies. NSL-KDD is the dataset suggested to solve some of the problems of the KDD 99 dataset. It removes all the redundant records in the KDD 99 train set, ensuring no duplicate records in the proposed test sets. NSL-KDD's KDD training set contains 22 attack types, and KDDTest data contains additional 17 attack types. It has 41 attributes and one class attribute, respectively. Another dataset commonly explored for this use case is The ISCX dataset 2012 by the Information Security Center of Excellence at the University of New Brunswick. It contains over 2 million traffic

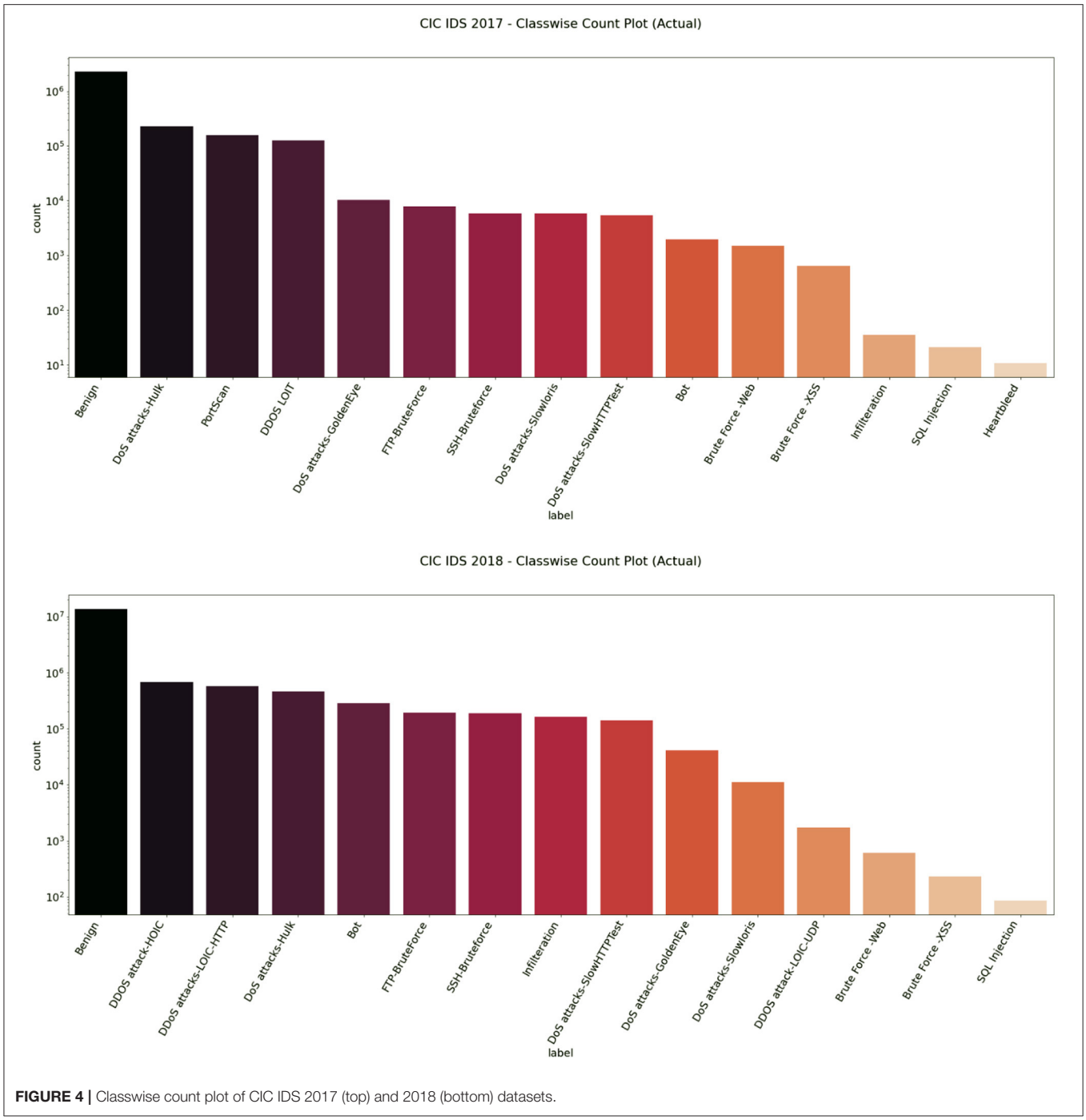


FIGURE 4 | Classwise count plot of CIC IDS 2017 (top) and 2018 (bottom) datasets.

packets categorized under 20 features. It covers attacks such as Brute Force SSH, HTTP DoS, DDoS and Infiltrating attacks. CIC analyzed these attacks under regular traffic rates on seven successive days, and they logged the corresponding metrics. It is one of the most up-to-date datasets.

CIC IDS 2017–2018 is the dataset we prefer for our Intrusion Detection System as it contains benign and the most up-to-date common attacks, which resembles the actual real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter with labeled flows based on the timestamp, source and destination IPs, source and destination ports, protocols, and attack. It is a labeled dataset containing a total number of 84 features, including their corresponding traffic status. Moreover, CIC IDS 2017 covers various attacks such as Brute Force Attack, HeartBleed Attack, Botnet, DoS Attack, Distributed DoS (DDoS) Attack, Web Attack, and Infiltration Attack. The total number of records in the dataset is 3,830,743. The benign traffic lists around 2,273,097 records, whereas the

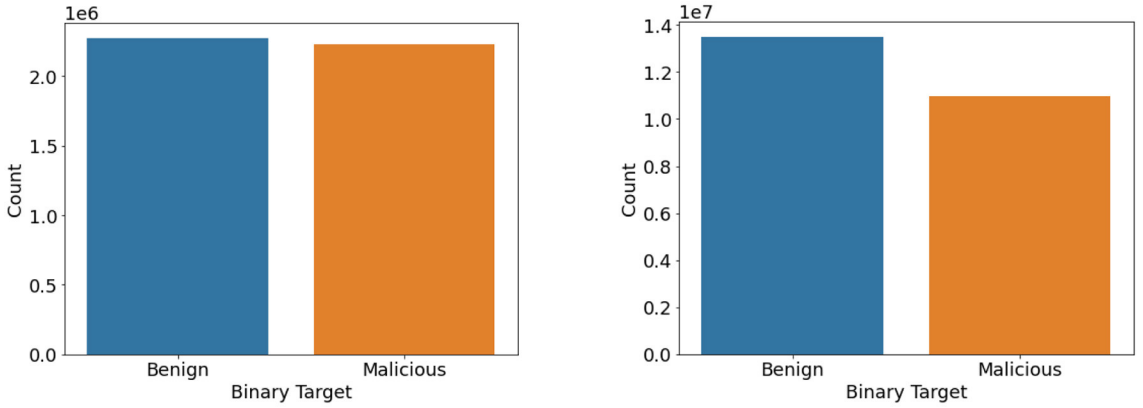


FIGURE 5 | Binary class count plot of oversampled CIC IDS 2017 (left) and 2018 (right) datasets.

malicious samples are 5,57,646 in number. The CIC IDS 2018 dataset consists of the same attack classes but with a large number of samples for benign and malicious classes. It consists of around 13,484,708 benign samples and 2,748,235 malicious samples.

Apart from this, we have also used the latest CIC Bell DNS 2021 dataset, which contains up-to-date real-time DNS-related data that can be useful for flagging a particular request as benign, spam, phishing, and malware. CIC created this dataset in collaboration with Bell Canada and Cyber Threat Intelligence, Canada. There were 32 features in the dataset, formed from DNS-statistical features and lexical features. The classes in the dataset were further grouped into 400,000 benign entries and 13,011 malicious entries, from which they balanced the classes by grouping 20,000 benign scans and 13,011 malicious scans together. **Table 1** displays the meta-data such as dataset size, number of entries, provider, and attack classes in detail for KDD Cup 99, CIC IDS 2017–2018, and CIC Bell DNS 2021.

Feature Selection

There were 82 features present in both CIC IDS 2017 and 2018 datasets. **Table 2** displays some of the features present in the dataset. Out of these, there were a lot of similar features that were highly related to each other, like the forward packet's total length and header length of packets. Deriving features with less correlation would help the algorithm learn diversely as they comparatively have lesser variance. Since both CIC IDS 2017 and 2018 had the same features, we calculated the correlation by combining both datasets as they belong to the same distribution.

We have employed Correlation-based Feature Selection (CFS) (37). By plotting the correlation matrix of the features (**Figure 2**), we analyzed which features were highly correlated to elucidate which ones were independent. Features with high correlation are more linearly dependent, and they tend to have the same effect on the dependent variable. So, when we observe two features with a relatively high correlation, we can drop one of those features, which reduces the computation complexity and dataset size to be processed. Around 56 features were derived that were correlated <75%. Further, we used recursive feature elimination (11, 37) by using a logistic regression classifier and selected 38 features based on their significance with an alpha value (p-value) of 0.05. **Table 2** represents some of the features in the dataset with corresponding correlation scores and p-value based significance.

Table 3 displays some of the selected features from the CIC Bell DNS 2021 dataset. The dataset was subject to multiple feature engineering techniques in processing the raw PCAP files into DNS statistical, lexical, and third-party features, as Mahdavifar et al. (8) suggested. We have used all 32 derived features from the dataset and have applied different preprocessing techniques discussed in section Data Preparation.

Improving Class Balance

Balanced datasets lead to good accuracy and performance of the classifiers. We found that the dataset had a heavy class imbalance by plotting the count of benign and intrusion samples. The number of benign samples was higher than the number of intrusion samples. One way to overcome this quickly would be to under-sample it. Since we have samples <100 for a few classes like SQL Injection in the CIC IDS 2018 dataset as referred to in **Figures 3**, **4**, under-sampling the imbalance dataset would still yield an imbalanced dataset with fewer samples. It may cause a bias toward higher-sample labels while training any algorithm.

We then used the Synthetic Minority Oversampling Technique algorithm to oversample the CIC IDS 2017 and 2018 datasets by a factor of 3 only for the intrusion class. Imbalance classification involves developing models on datasets that have a severe class imbalance. Synthetic Minority Oversampling Technique or SMOTE algorithm is a data augmentation technique that selects features in close feature space. It creates as many synthetic examples for minority classes, making them more effective in learning the decision boundary and performing better. SMOTE uses K-Nearest Neighbors to select points near the neighbor of each data point randomly and generates synthetic instances to upsample the minority class. SMOTE oversampling is effective because the new synthetic samples generated from the minority class have relatively close feature space to existing samples. In this way, we were able to get an almost balanced dataset, leading to better performance of the model. **Figures 5**, **6**

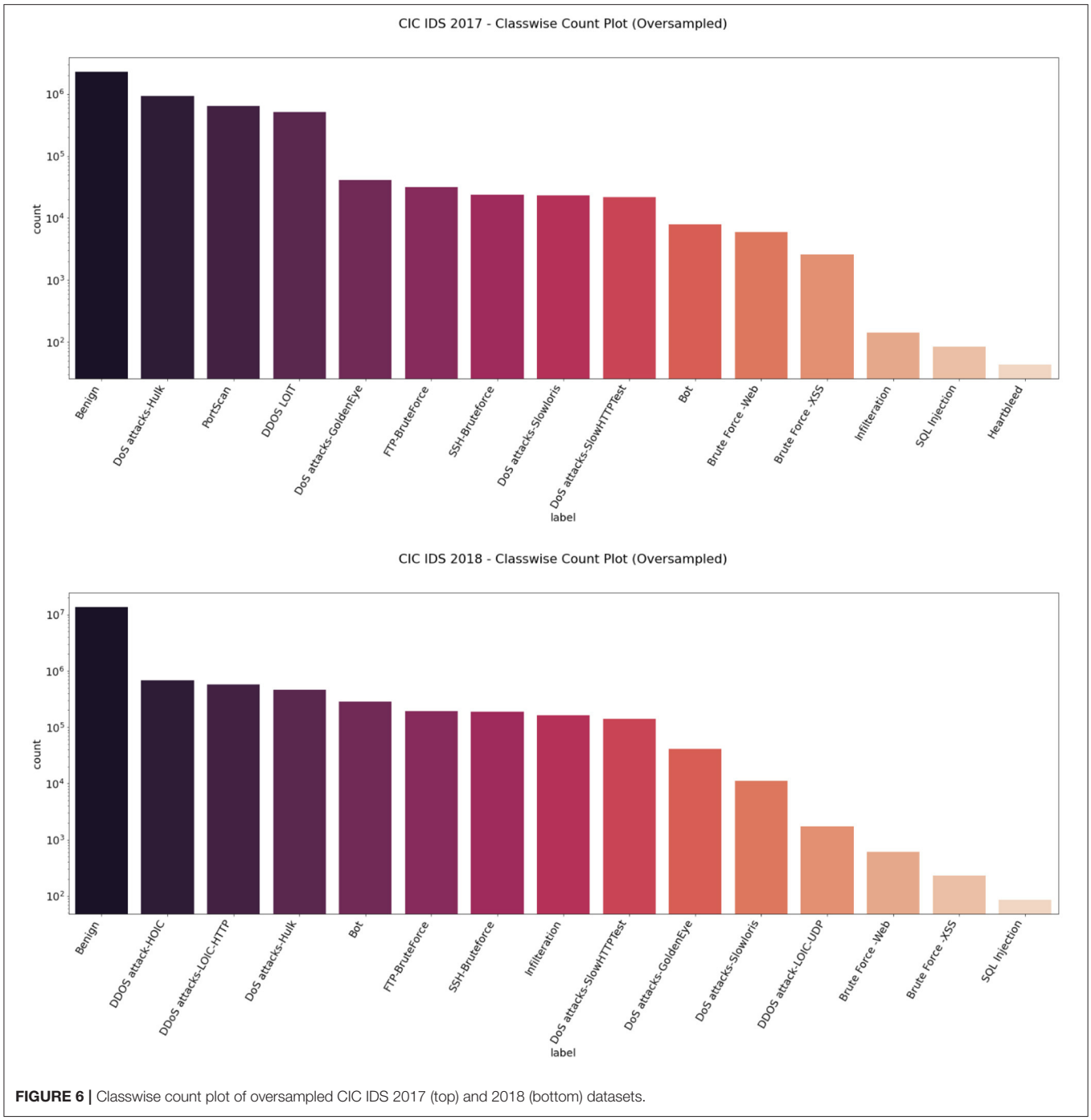


FIGURE 6 | Classwise count plot of oversampled CIC IDS 2017 (top) and 2018 (bottom) datasets.

represent the oversampled binary count plot and classwise count plot of the CIC IDS 2017 and 2018 datasets.

We performed both undersampling and oversampling for the CIC Bell DNS 2021 to promote class balance. The dataset had a heavy class imbalance with 400,000 benign samples and only 13,011 malicious samples. Random shuffling and selection were performed on the dataset to gather only around 25,000 benign samples. Apart from this, SMOTE oversampling was done on the rest of the three classes: spam, phishing, malware. We also removed entries that contained null values since they were not usable. After this process, we had a balanced dataset with 23,716 benign samples and 22,929 malicious samples. **Figures 7**, **8** illustrate the binary and classwise count plots of samples in the CIC Bell DNS 2021 dataset before and after the class balance improvement process.

Table 4 shows the number of samples and benign to malicious sample ratio before and after class balance improvement for CIC IDS 2017, CIC IDS 2018, and CIC Bell DNS 2021 datasets.

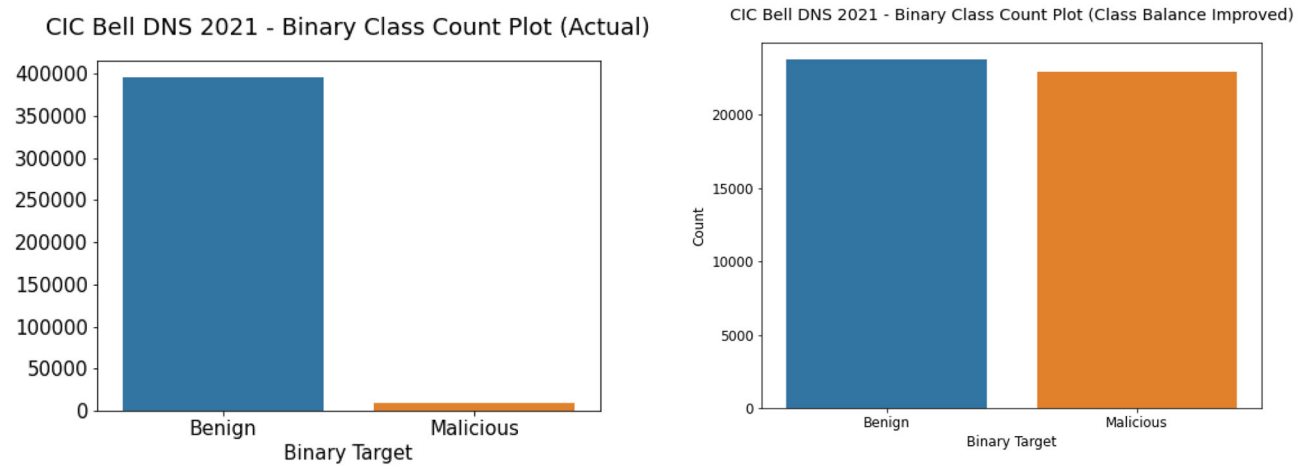


FIGURE 7 | Binary class count plot of actual (left) and class balance improved (right) CIC Bell DNS 2021 dataset.

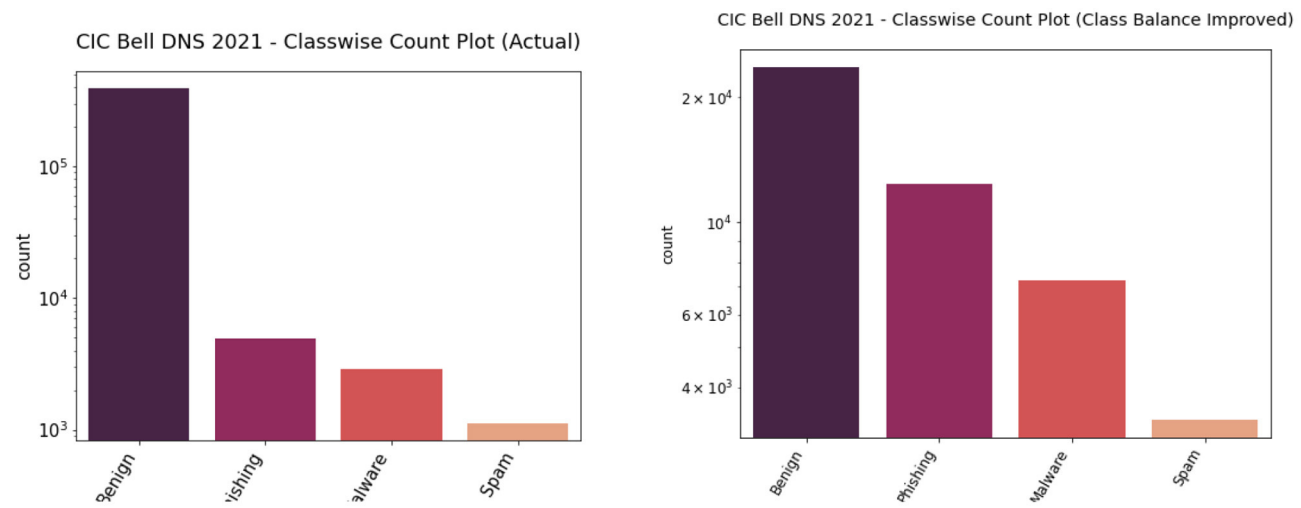


FIGURE 8 | Classwise count plot of actual (left) and class balance improved (right) CIC Bell DNS 2021 dataset.

TABLE 4 | Data sample numbers and percentage before and after oversampling CIC IDS 2017, 2018, and CIC Bell DNS 2021 datasets.

	Actual		Balanced		Actual Percentage		Balanced Percentage	
	Benign	**Malicious**	**Benign**	**Malicious**	**Benign**	**Malicious**	**Benign**	**Malicious**
CIC IDS 2017	2,273,097	557,646	2,273,097	2,230,584	80.3	19.7	50.47	49.53
CIC IDS 2018	13,484,708	2,748,235	13,484,708	10,992,940	83.07	16.93	55.09	44.91
CIC Bell DNS 2021	400,000	13,011	23,716	22,929	96.84	3.16	50.84	49.16

Data Preparation

The dataset was made publicly available by the Canadian Institute for Cybersecurity (CIC). The CIC Bell DNS 2021 dataset's processed CSV files were downloaded from their website. We performed standard data cleaning to remove null values and encode categorical text such as domain names, countries, and states with numbers. The CIC IDS 2017 and 2018 datasets were downloaded from their AWS S3 bucket using the AWS CLI and split into multiple CSV files based on the type of attack. We then combined these CSV files as a single data frame during runtime to save memory for further preprocessing as the entire dataset was more than 7 GB. The dataset was subject to multiple feature selection and sampling processes as mentioned in the previous sections to promote the most valuable features for classification

to achieve better accuracy. We combined the attack columns into three feature columns: labels as the attack's name, binary targets denoting whether the packet is benign or malicious, and target column representing the id of which attack had occurred. Later this data was z-scored using the standardization formula (1). Standardization of the data leads to faster and more accurate convergence of the machine learning models.

$$y = \frac{(x - \bar{x})}{\sigma}$$

We followed a stratified cross-validation split with a 30% test size on all three datasets while running our algorithmic experiments. The cross-validation split yielded a 70% training set, 15% validation set, and 15% testing set for every fold with a random shuffling of the data entries. Since the split was stratified, we contained the same benign to malicious samples ratio in all three sets for each dataset.

Machine Learning Approaches

We experimented with five machine learning algorithms inspired by different pieces of contemporary machine intelligence literature whose methods proved to work well for tabular data and critical applications. Logistic Regression is a statistical learning algorithm used to model the probability of an event happening or not. The logistic function or sigmoid function approximates the output values from a multinomial linear equation to predict whether a particular class is occurring. We used L2 regularization and "lbfgs" solver to optimize the algorithm. Equation 2 defines the logit calculation by summation over the dot product of inputs and weights added to bias, which is then applied inside sigmoid, equation 3, to get a probability-like value between 0 and 1, which denotes the occurrence of a class.

$$z = w^T x + b$$

$$\sigma(z) = \frac{1}{1 + e^{-(z)}}$$

Binary Cross entropy Loss

$$= -\frac{1}{output\ size} \sum_{i=1}^{output\ size} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i)$$

Binary Cross Entropy or BCE (4) is used as the cost function with L2 regularization for the optimization process using gradient descent.

Further, we used Decision trees with the ID3 implementation and Gini criterion. We also ran Random forests with 100 trees and criteria as Gini, equation 5. This algorithm establishes the outcome based on the predictions of the decision trees, which uses the information gain, equation 6, as an optimization metric.

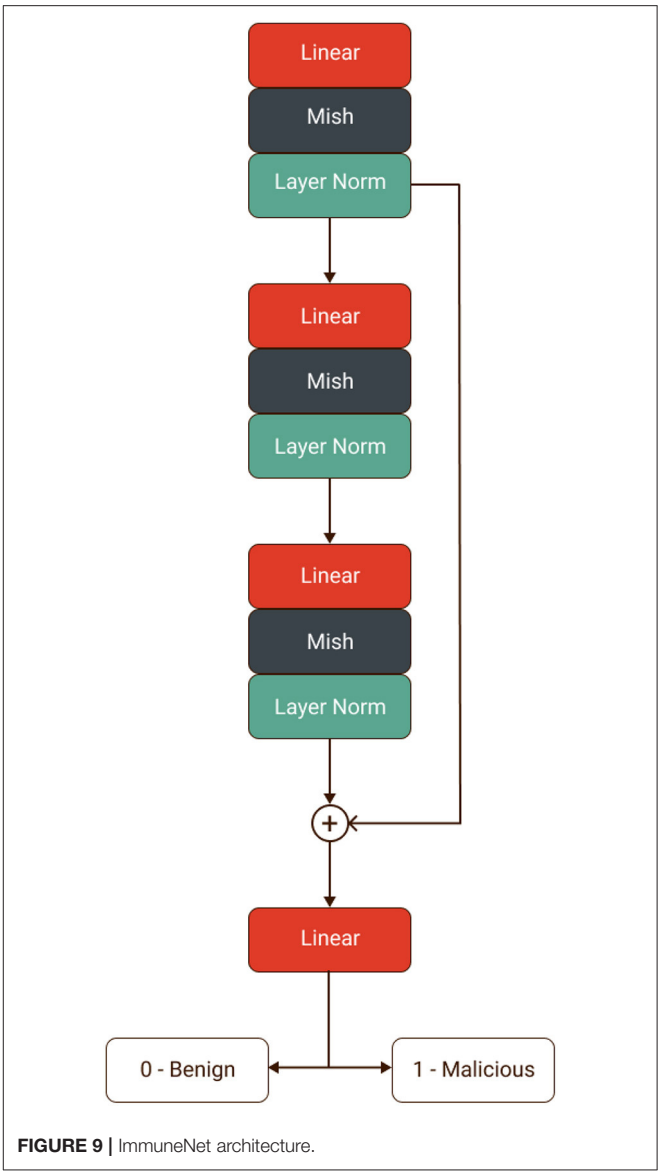


FIGURE 9 | ImmuneNet architecture.

$$Gini: I_G(f) = \sum_{i=1}^{m} f_i \ (1 - f_i)$$

$$Information\ gain: I_E(f) = \sum_{i=1}^{m} f_i \ \log_2 \quad f_i$$

Extreme Gradient Boosted Trees (XGB) (12) have recently gained momentum for their outstanding results on different kinds of tabular data. With regularization techniques and effective pruning of the trees, only significant splits with positive information gain are retained. We observed that the above methods, namely Extreme Gradient Boosted Trees, Random Forests, and Decision Trees, worked better with given use cases than Logistic Regression because ensemble models use

TABLE 5 | Hyperparameter details for ImmuneNet for CIC IDS 2017, CIC IDS 2018, and CIC Bell DNS 2021 datasets.

Hyperparameters	Values as per dataset		
	CIC IDS 2017	**CIC IDS 2018**	**CIC bell DNS 2021**
Input feature dimension	38	38	32
Batch size	64	64	64
Adam optimizer	$\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon =$ 1e-7	$\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon =$ 1e-7	$\beta_1 = 0.9$, $\beta_2 = 0.99$, $\varepsilon =$ 1e-7
Learning rate	1e-4	1e-4	3e-4
Additional optimization strategies	Cosine annealing learning rate scheduling, early stopping	Cosine annealing learning rate scheduling, early stopping	Cosine annealing learning rate scheduling, early stopping
# Epochs	15	15	15
# Units in input FC layer	512	512	512
Input FC layer activation	Mish	Mish	Mish
# Units in hidden FC layer 1	784	784	784
FC layer 1 activation	Mish	Mish	Mish
# Units in hidden FC layer 2	512	512	512
FC layer 2 activation	Mish	Mish	Mish
# Units in output FC layer (binary classification)	1	1	1
Output FC layer activation	Sigmoid	Sigmoid	Sigmoid

TABLE 6 | Number of samples in training, validation, and testing set for CIC IDS 2017–2018, and CIC Bell DNS 2021 datasets.

	Number of samples in the training set	**Number of samples in the validation set**	**Number of samples in the test set**	**Total**
CIC IDS 2017	3,152,567	675,552	675,552	4,503,681
CIC IDS 2018	17,134,354	3,671,647	3,671,647	24,477,648
CIC Bell DNS 2021	32,652	6,996	6,997	46,645

multiple independent learners from which the entire predictions are derived.

Finally, we built our own Artificial Neural Network named the "ImmuneNet" that uses simple linear projections followed by mish activation and layer normalization with a residual operation between the first and the final block. **Figure 9** displays the neural network architecture of ImmuneNet.

The first and last linear layer has 512 neurons, and the hidden layer has 784 neurons. Since there were fewer features for all three datasets, we used a minimal number of neurons in the layers. The number of neurons per layer was set arbitrarily and increased by a factor of 2 after observing the performance of the earliest models by using a grid search with different hyper-parameter settings for the hidden layer units. The first and the last layer are connected with a residual (38) operation, and hence they have the same number of neurons. Equation (7) defines the perceptron or linear equation, which is chained with the mish (25) activation (8) followed by the layer normalization (39) (10), creating a forward block for the immune net. We have used Sharma and Yadav (10) technique for the neural network's weights initialization for training all three datasets. **Table 5** describes the hyperparameters we used for ImmuneNet and its optimization for all three datasets with corresponding feature dimensions. The number of samples used for training, validating, and testing the algorithms is mentioned below in **Table 6**.

We used the Adam optimization (40) set to a learning rate of 1e-4 with cosine annealing (36) technique for learning rate scheduling and binary cross-entropy (4) as the cost function. The hyper-parameters for the optimizer except for the learning rate, like beta1, and beta2 have been set with values as suggested by the authors of Adam optimizer (40). The learning rate will be changed after every few iterations based on the validation loss obtained, with the help of the cosine annealing (36) learning rate scheduler. The cosine function is used to vary the lower rate after each epoch. The objective of cosine annealing is to prevent the model from saturating after reaching a local minima and varying the learning rate using the cosine function to periodic lower learning rates to converge at an optimal point like the global minima. The model was trained with a batch size of 64 for 15 epochs with an early stopping mechanism which stops the training process when the model's performance starts to saturate, which acts as a regularization technique. We have avoided larger batch sizes because models with larger batch sizes and higher learning rates can cause unstable training, bad convergence, and lead to different asymptomatic test metrics. The results of all

TABLE 7 | Confusion matrix for binary classification.

	Class 0	**Class 1**
Class 0	TP	FP
Class 1	FN	TN

TABLE 8 | Comparison of models based on respective evaluation metrics for CIC IDS 2017.

Model	Accuracy	Precision	Recall	F1	ROC-AUC Score	Training Loss	Validation Loss	Testing Loss
ImmuneNet	99.63	99.64	99.63	99.63	99.636	0.0077	0.0104	0.0091
XGB	99.09	99.03	99.07	99.07	99.10	0.0150	0.0239	0.0216
Random forest	98.81	98.82	98.81	98.81	98.31	0.0388	0.0272	0.274
Decision trees	98.54	93.41	96.79	95.06	96.79	0.1132	0.0986	0.0961
Logistic regression	92.96	90.8	90.96	90.87	91.5	0.2341	0.2560	0.2663

the algorithms mentioned above are discussed in detail under section Results.

Linear Layer:

$$y = w^T.x+b$$

Mish Activation:

$$f(x) = x\tanh(\ln(1+e^x))$$

Layer Normalization

$$\hat{x}_{i,l} = \frac{x_{i,k}-\mu_i}{\sqrt{\sigma_i^2+\epsilon}}$$
$$y_i+\gamma\hat{x}_i+\beta \equiv LN_{\gamma,\beta}(x_i)$$

RESULTS

Evaluation Metrics

The following metrics for evaluating our machine learning models can help derive more significant insights into our various algorithms' performance and bias:

Confusion Matrix (**Table 7**) is a matrix that summarizes the performance of the given algorithm on the classification. Since the dataset is not entirely balanced, using measures like accuracy alone can be misleading.

Accuracy, equation (11), or correct rate is how right the model is for the given classification problem.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision (12) is the ratio of true positives to all the positives, i.e., the sum of true positives and false positives.

$$Precision = \frac{TP}{TP+FP}$$

Recall (13) is the ratio between true positives and the sum of true positives and false negatives.

$$Recall = \frac{TP}{TP+FN}$$

The F1 (14) score provides the balance between the precision and recall, suggesting how similar the predicted set is to the true set.

$$F1 = \frac{2^*\text{Precision}^*\text{Recall}}{Precision+Recall} = \frac{2^*\text{TP}}{2^*\text{TP}+FP+FN}$$

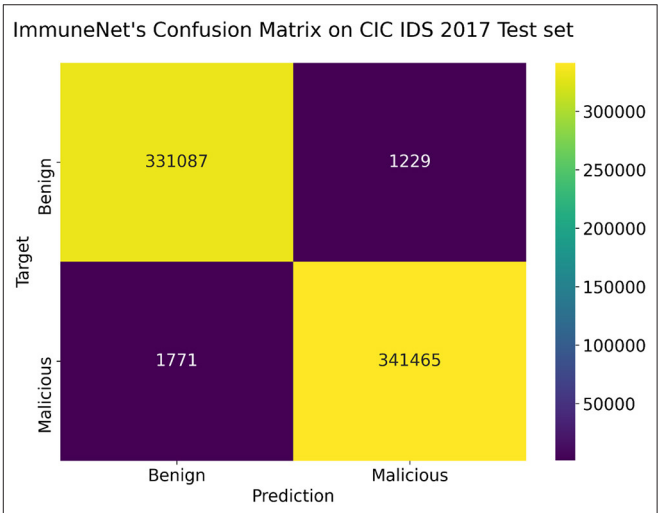


FIGURE 10 | Confusion matrix on the CIC IDS 2017 test set for ImmuneNet.

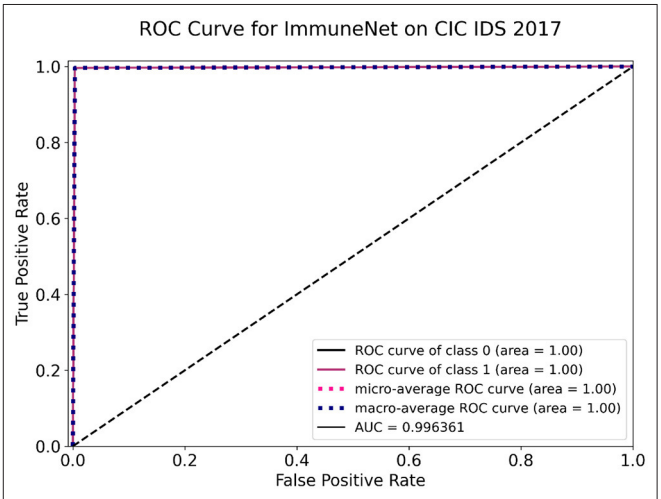


FIGURE 11 | Receiver operating characteristic (ROC) curve of ImmuneNet on CIC IDS 2017 dataset.

ROC Curve is the rate of true positives against false negatives describing how well a binary classifier can classify between two classes. The area of the region under the ROC curve defines the performance of the binary classifier. The model performs better when it has a high AUC score (15) and performs the lowest when

TABLE 9 | Comparison of models based on respective evaluation metrics for CIC IDS 2018.

Model	Accuracy	Precision	Recall	F1	ROC-AUC Score	Training loss	validation loss	Testing loss
ImmuneNet	99.78	99.77	99.78	99.7	99.786	0.0036	0.0023	0.0025
XGB	99.00	99.03	99.00	99.01	98.98	0.0141	0.0366	0.0389
Random forest	98.81	98.82	98.81	98.81	98.31	0.0989	0.0782	0.0801
Decision trees	98.69	93.41	98.79	96.03	98.731	0.1054	0.1031	0.1053
Logistic regression	87.96	90.8	87.96	88.99	81.537	0.3520	0.2877	0.2798

the AUC score is low.

$$AUC = \frac{Specificity + Sensitivity}{2}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

Binary Crossentropy (3) is a cost function that gives the log distance between two probability distributions containing probabilities of an event occurring or not for any binary classification problem.

$$Binary\ Crossentropy\ Loss = \frac{1}{output\ size} \sum_{i=1}^{output\ size} y_i \cdot \log \hat{y_i} + (1 - y_i) \cdot \log(1 - \hat{y_i})$$

Model Comparison

The above models were tested on the test sets of CIC IDS 2017, CIC IDS 2018, and CIC IDS Bell DNS 2021 datasets, as mentioned in section Improving Class Balance. **Table 8** contains the evaluation metrics of the algorithms' performance on the CIC IDS 2017 dataset. **Figures 10**, **11** display the confusion matrix and ROC curve of ImmuneNet on the CIC IDS 2017's test set.

From **Table 8**, we can observe that the Immune Net algorithm performs the best, with an accuracy of 99.63%. Compared to the other algorithms like XGB (12) and Random Forests, ImmuneNet has better accuracy and minimum loss. From the Confusion Matrix represented in **Figure 10**, we can also see that only 1,229 benign samples were flagged as malicious in total, implying a less false positive rate. On the other hand, XGB (12) and Random Forest perform better when compared to Decision Trees and Logistic Regression with an accuracy of 99.09 and 98.81 %.

The Receiver Operating Characteristic curve for ImmuneNet, as seen in **Figure 11**, has an AUC score of 99.6% for the CIC DIS 2017 test set. This makes it evident that the algorithm's true positive rate against the false positive rate is better and is almost a typically perfect binary classifier.

Table 9 comprises the evaluation metrics of the five machine learning algorithms on the CIC IDS 2018 dataset. The confusion matrix and ROC curve of ImmuneNet are represented by **Figures 12**, **13** for the CIC IDS 2018's test set.

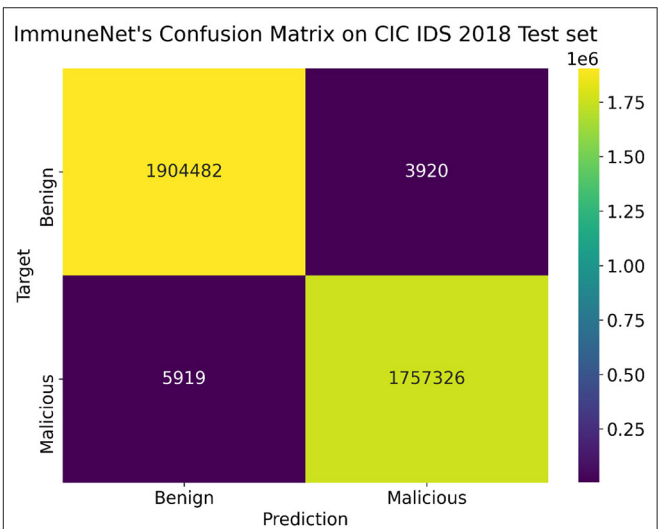


FIGURE 12 | Confusion matrix on the CIC IDS 2018 test set for ImmuneNet.

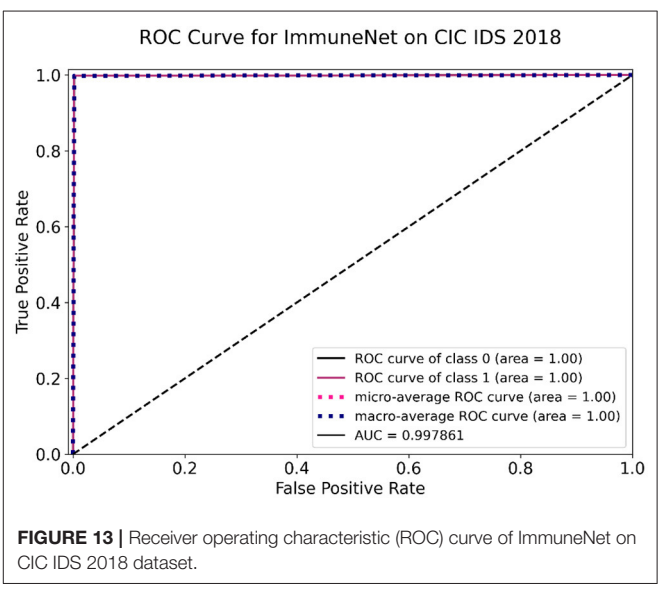


FIGURE 13 | Receiver operating characteristic (ROC) curve of ImmuneNet on CIC IDS 2018 dataset.

ImmuneNet has an accuracy of 99.78, 99.77% precision score, 99.78% recall score, and a test loss of 0.0025, which is better when compared to the other algorithms. Logistic Regression had the lowest accuracy of 87.96% and had a high loss. **Figure 12** depicts

TABLE 10 | Comparison of models based on respective evaluation metrics for CIC Bell DNS 2021.

Model	Accuracy	Precision	Recall	F1	ROC-AUC Score	Training Loss	Validation Loss	Testing Loss
ImmuneNet	99.19	99.22	99.19	99.20	99.199	0.0124	0.0143	0.00141
XGB	99.10	98.89	99.10	98.99	99.01	0.0235	0.0228	0.0226
Random forest	98.21	95.60	98.17	96.86	96.25	0.0789	0.0631	0.0655
Decision trees	97.66	95.41	98.78	97.06	96.30	0.0808	0.0748	0.0752
Logistic regression	92.87	89.93	92.60	91.24	90.10	0.3356	0.3579	0.3467

the Confusion Matrix, from which we can observe that the false positives and false negatives are very low, having only samples around 3,920 and 5,919, respectively.

From the ROC curve, as presented in **Figure 13**, we can observe that ImmuneNet's performance has a better True Positive rate against False Positive rate producing a ROC-AUC score of 99.78%. The obtained metrics are higher than other existing methods for CIC IDS 2018, such as the LighGBM and HBGB method proposed by Seth et al. (19).

Several evaluation metrics in **Table 10** depict the algorithms' performance on the CIC Bell DNS 2021 dataset. ImmuneNet's confusion matrix and ROC curve on the CIC Bell DNS 2021's test set are represented by **Figures 14**, **15**.

From **Table 10**, it is observed that the ImmuneNet has the highest accuracy of about 99.2% and a test loss of 0.00141. These metrics are much better when compared with the algorithms like Decision Trees, Random Forest and Logistic Regression, which have comparatively lower accuracies and higher losses.

From **Figure 14**, we can see that ImmuneNet has an excellent true positive rate against false positive rate, having a ROC-AUC score of 99.19%. It is evident that ImmuneNet is performing better on the CIC Bell DNS 2021 dataset from the metrics mentioned above compared to Mahdavifar et al. (8). They achieved about 94.8 and 99.4% accuracy on the 60/40% balanced dataset and the 97/3% balanced dataset. The discussion on the performance of the algorithms mentioned above for all three datasets with their advantages and limitations is elaborated under the next section.

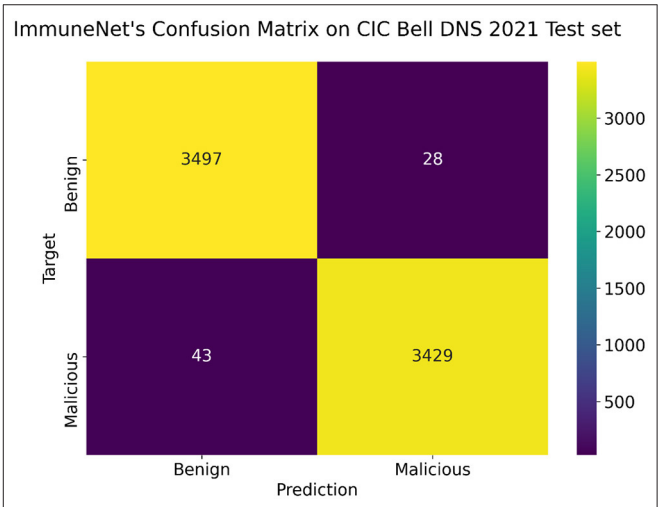


FIGURE 14 | Confusion MATRIX on the CIC Bell DNS 2021 test set for ImmuneNet.

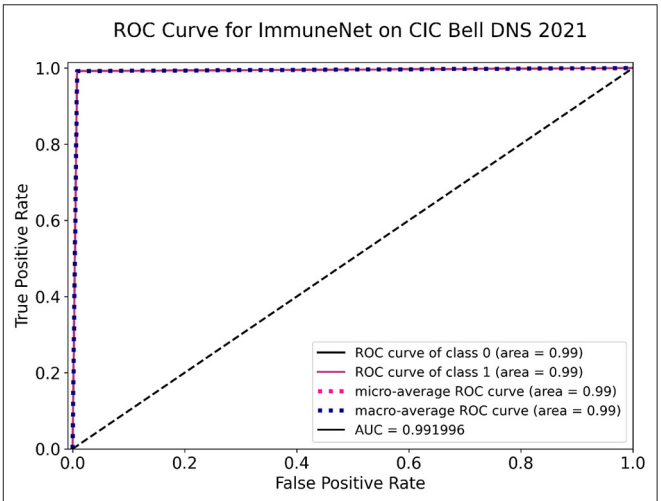


FIGURE 15 | Receiver operating characteristic (ROC) curve of ImmuneNet on CIC Bell DNS 2021 dataset.

DISCUSSION

Out of all the experiments run on the CIC IDS 2017 (6), CIC IDS 2018 (7), and CIC Bell DNS 2021 (8) datasets with over five types of machine learning algorithms, ImmuneNet performed the best for all three datasets. This is because the algorithms' training, validation, and testing phases were data-centric and highly dependent on the class balancing and feature engineering techniques we followed for the three datasets. ImmuneNet achieved ∼99.2% accuracy for the CIC Bell DNS 2021 dataset, 99.8% accuracy for CIC IDS 2018 dataset, and 99.63% accuracy for the CIC IDS 2017 dataset during the testing phase. These metrics are better than the existing approaches that have been suggested for IDS in healthcare, such as methods proposed by Subasi et al. (4), Šabić et al. (15), Hady et al. (16) since the model has been trained on an extensive real-time dataset containing latest cyberattacks post multiple feature engineering and sampling techniques. The algorithm has also performed well-after performing a different sampling and preprocessing approach for the CIC Bell DNS 2021 dataset, achieving an

TABLE 11 | ImmuneNet's model parameters and size for CIC IDS 2017–2018, and CIC Bell DNS 2021 datasets.

S. No	Dataset	Number of trainable parameters	Model weights size (in megabytes)
1	CIC IDS 2017	828,209	3.265
2	CIC IDS 2018	828,209	3.265
3	CIC Bell DNS 2021	825,137	3.253

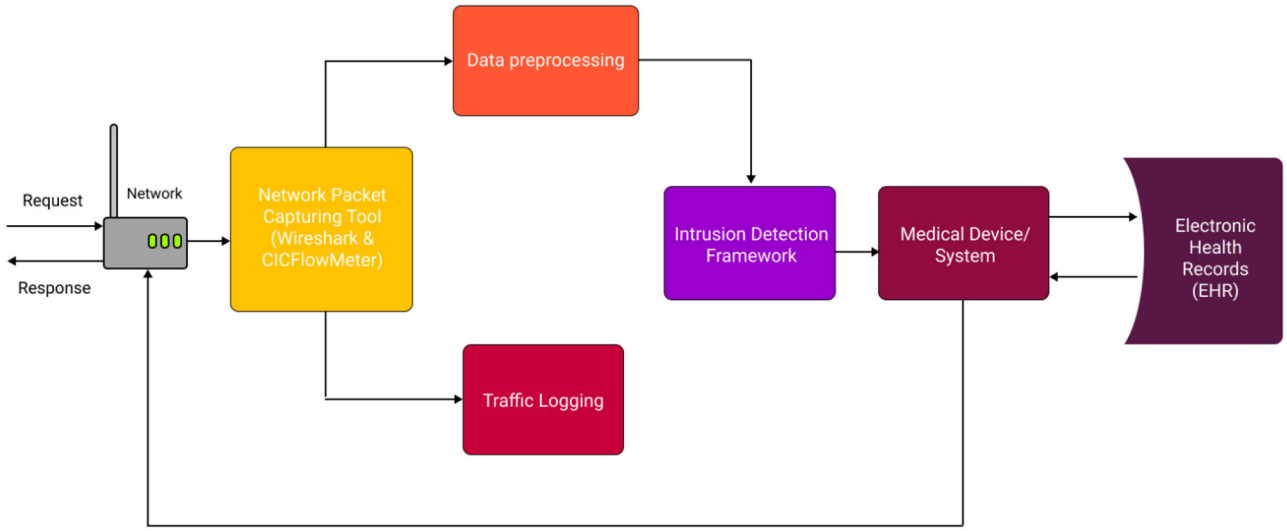


FIGURE 16 | System architecture diagram for the proposed intrusion detection system.

accuracy of about 99.2%, which is comparatively better than the method proposed by Samaneh Mahdavifar et al. (8).

The four reasons behind ImmuneNet's good performance are the data preprocessing, residual operation (38), mish activation (25), and layer normalization (39). By adding the tensors from the first layer and the tensors from the last layer, we preserve features learned by the first layer even if the features learned by the last layer were terrible. This also makes a difference during backward propagation. Mish activation (25) is unbounded above, allowing positive values and slight negative values to flow through the network without any saturation caused by capping during the forward and backward propagation. Layer normalization (39) normalizes the input tensor to zero mean and unit variance along the feature direction. The feature selection, synthetic sampling techniques, and data standardization during preprocessing have led to well-balanced and processed data from which the model was able to learn better.

Apart from ImmuneNet, other algorithms like XGB (12) and Random Forest also performed competitively, having the best subsequent metrics for all three datasets. The learning algorithms must be trained on balanced datasets to achieve better performance and unbiased detection of malicious requests. The CIC Bell DNS 2021 had a heavy class imbalance, and the number of samples obtained after the data preparation was low compared to CIC IDS 2017 (6) and 2018 (7) datasets. The CIC Bell DNS dataset can be improved by including more real-time samples for the phishing, spam, and malware domains to make it more balanced, and the algorithm can be fine-tuned to improve its performance. The ImmuneNet model is very lightweight, containing <1 million parameters. **Table 11** displays the parameters and size of the model's weight in megabytes for all three architectures. This makes it easy to deploy it on any medical device or healthcare system without occupying much memory and would also be fast during inference and fine-tuning. However, the model has to be fine-tuned to support new types of attacks in the future to keep it abreast and safeguard healthcare systems with high accuracy. The neural network's weights can be frozen except the final classification layer, and the framework can be trained again with new data to fine-tune the model. The training process can be done again with new data to retrain the model and improve its performance. **Figure 16** displays the system architecture of the proposed intrusion detection system showing the typical deployment setting and entire process flow.

CONCLUSION

The proposed neural network model, ImmuneNet, has given the highest accuracy of over 99.2% on the class balance improved CIC Bell DNS 2021 dataset, 99.8 and 99.63% accuracy on the oversampled CIC IDS 2018 and 2017 datasets. The model obtained a ROC-AUC score of 99.19%, which is considerably better than the score achieved by the other approaches. We followed a data-centric approach and used several feature

selection and sampling techniques to improve class balance and the pre-processing of the datasets, leading to these results. This architecture is the latest in terms of Intrusion Detection as it has been trained with three latest real-time and extensive datasets. It is deployment-friendly for medical devices and healthcare systems because of its lightweight and fast characteristics. This would help protect the patient records from any kind of cyber attack, leading to consistency in patients' medical data. It is also highly critical that the models are not biased toward specific attacks since bias can allow certain attacks to bypass the IDS and manipulate the patient records by entering into the system. ImmuneNet has an excellent true positive rate against false positive rate, implying it is not biased toward false positives and negatives. However, the CIC DNS 2021 dataset needs to be improved with more samples from phishing, spam, and malware domains to promote natural class balance and better benchmarking of the algorithms. The proposed system must also be frequently fine-tuned with new cyber attack data to protect patient records from cyber attacks that emerge in the future. Fine-tuning the model on the latest cyber-attack data from time to time would improve the model's performance and keep it up-to-date, but it is a time-consuming and tedious process. To overcome these disadvantages, our future work would include a self-supervised learning-based neural network architecture that can make systems like intrusion detection adapt by themselves to new attacks without any manual retraining or fine-tuning and focusing on the recent datasets with better class balance and adversarial reweighting strategies to improve the fairness of the model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

C-YC did the conceptualization, supervised the research, and carried out the funding acquisition. MA and DS investigated the data, performed the methodology, and implemented the software code. KS and C-YC carried out the project administration, validated the results, reviewed, and edited the manuscript. MA, DS, PV, and HG wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partially funded by Intelligent Recognition Industry Service Research Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and Ministry of Science and Technology in Taiwan (Grant No. MOST 109-2221-E-224-048-MY2).

REFERENCES

1. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm. *J Circ Syst Comp.* (2021) 15:2250010. doi: 10.1142/S0218126622500104
2. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M. The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors.* (2020) 20:2559. doi: 10.3390/s20092559
3. Yeng PK, Nweke LO, Woldaregay AZ, Yang B, Snekkenes EA. Data-driven and artificial intelligence (AI) approach for modelling and analyzing healthcare security practice: a systematic review. In: Arai K, Kapoor S, Bhatia R, editors. *Intelligent Systems and Applications*. Cham: Springer (2021). doi: 10.1007/978-3-030-55180-3_1
4. Subasi A, Algebsani S, Alghamdi W, Kremic E, Almaasrani J, Abdulaziz N. Intrusion detection in smart healthcare using bagging ensemble classifier. In: *International Conference on Medical and Biological Engineering*. Cham: Springer (2021) p. 164–71. doi: 10.1007/978-3-030-73909-6_18
5. Sarna Priya RM, Maddikunta PK, Parimala M, Koppu S, Gadekallu TR, Chowdhary CL, et al. An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. *Comp Commun.* (2020) 160:139–49. doi: 10.1016/j.comcom.2020.05.048
6. CIC IDS. *Dataset*. (2017). Available online at: https://www.unb.ca/cic/datasets/ids-2017.html
7. CIC IDS. *Dataset.* (2018). Available online at: https://registry.opendata.aws/cse-cic-ids2018/
8. Mahdavifar S, Maleki N, Lashkari AH, Broda M, Razavi AH. Classifying malicious domains using DNS traffic analysis. In: *The 19th IEEE International Conference on Dependable, Autonomic, and Secure Computing (DASC)*. Calgary, AB (2021).
9. Gopalan SS, Ravikumar D, Linekar D, Raza A, Hasib M. Balancing approaches towards ML for IDS: a survey for the CSE-CIC IDS dataset. In: *2020 International Conference on Communications, Signal Processing, and Their Applications (ICCSPA)*. Sharjah: IEEE (2021). doi: 10.1109/ICCSPA49915.2021.9385742
10. Sharma NV, Yadav NS. An optimal intrusion detection system using recursive feature elimination and ensemble of classifiers. *Microproc Microsyst.* (2021) 23:104293. doi: 10.1016/j.micpro.2021.104293
11. Singh K, Mahajan A, Mansotra V. Using recursive feature elimination and fisher score with convolutional neural network for identifying port scan attempts. In: *Smart Trends in Computing and Communications*. Singapore: Springer (2022). p. 551–60. doi: 10.1007/978-981-16-4016-2_52
12. Tandon A, Sharma R, Sodhiya S, Vincent PM. QR code based secure OTP distribution scheme for authentication in net-banking. *Int J Eng Technol.* (2013) 5:0975–4024.
13. Tervoort T, De Oliveira MT, Pieters W, Van Gelder P, Olabarriaga SD, Marquering H. Solutions for mitigating cybersecurity risks caused by legacy software in medical devices: a scoping review. *IEEE Access.* (2020) 8:84352–61. doi: 10.1109/ACCESS.2020.2984376
14. Thamilarasu G, Odesile A, Hoang A. An intrusion detection system for internet of medical things. *IEEE Access.* (2020) 8:181560–76. doi: 10.1109/ACCESS.2020.3026260
15. Šabić E, Keeley D, Henderson B, Nannemann S. Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data. *AI Soc.* (2021) 36:149–58. doi: 10.1007/s00146-020-00985-1
16. Hady AA, Ghubaish A, Salman T, Unal D, Jain R. Intrusion detection system for healthcare systems using medical and network data: a comparison study. *IEEE Access.* (2020) 8:106576–84. doi: 10.1109/ACCESS.2020.3000421
17. Nguyen PT, Huynh VD, Vo KD, Phan PT, Elhoseny M, Le DN. Deep learning based optimal multimodal fusion framework for intrusion detection systems for healthcare data. *CMC.* (2021) 66:2555–71. doi: 10.32604/cmc.2021.012941

18. Iwendi C, Anajemba JH, Biamba C, Ngabo D. Security of things intrusion detection system for smart healthcare. *Electronics.* (2021) 10:1375. doi: 10.3390/electronics10121375
19. Seth S, Chahal KK, Singh G. A novel ensemble framework for an intelligent intrusion detection system. *IEEE Access.* (2021) 9:138451–67. doi: 10.1109/ACCESS.2021.3116219
20. Serinelli BM, Collen A, Nijdam NA. On the analysis of open source datasets: validating IDS implementation for well-known and zero day attack detection. *Proc Comp Sci.* (2021) 191:192–9. doi: 10.1016/j.procs.2021.07.024
21. Thilagam T, Aruna R. Intrusion detection for network based cloud computing by custom RC-NN and optimization. *ICT Express.* (2021) 7:512–20. doi: 10.1016/j.icte.2021.04.006
22. Zuech R, Hancock J, Khoshgoftaar TM. Detecting web attacks using random undersampling and ensemble learners. *J Big Data.* (2021) 8:1–20. doi: 10.1186/s40537-021-00460-8
23. Yu L, Dong J, Chen L, Li M, Xu B, Li Z, et al. PBCNN: packet bytes-based convolutional neural network for network intrusion detection. *Comp Netw.* (2021) 194:108117. doi: 10.1016/j.comnet.2021.108117
24. Fitni QR, Ramli K. Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems. In: *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE (2020) p. 118–24. doi: 10.1109/IAICT50021.2020.9172014
25. Misra D. Mish: a self regularized non-monotonic neural activation function. *arXiv Preprint*. (2019) arXiv:1908.08681.
26. Hua Y. An efficient traffic classification scheme using embedded feature selection and lightgbm. In: *2020 Information Communication Technologies Conference (ICTC)*. Nanjing: IEEE (2020). doi: 10.1109/ICTC49638.2020.9123302
27. Catillo M, Rak M, Villano U. 2L-ZED-IDS: a two-level anomaly detector for multiple attack classes. In: *AINA Workshops*. Caserta: Springer (2020). p. 687–96. doi: 10.1007/978-3-030-44038-1_63
28. Khan MA, Kim J. Toward developing efficient Conv-AE-based intrusion detection system using heterogeneous dataset. *Electronics.* (2020) 9:1771. doi: 10.3390/electronics9111771
29. Meryem A, Ouahidi BE. Hybrid intrusion detection system using machine learning. *Netw Sec.* (2020) 2020:8–19. doi: 10.1016/S1353-4858(20)30056-8
30. Thakkar A, Lohiya R. A review of the advancement in intrusion detection datasets. *Proc Comp Sci.* (2020) 167:636–45. doi: 10.1016/j.procs.2020.03.330
31. Gao X, Shan C, Hu C, Niu Z, Liu Z. An adaptive ensemble machine learning model for intrusion detection. *IEEE Access.* (2019) 7:82512–21. doi: 10.1109/ACCESS.2019.2923640
32. Chawla A, Lee B, Fallon S, Jacob P. Host based intrusion detection system with combined CNN/RNN model. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Cham: Springer (2018). p. 149–58. doi: 10.1007/978-3-030-13453-2_12
33. Freas CB, Harrison RW, Long Y. High performance attack estimation in large-scale network flows. In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE (2018). p. 5014–20. doi: 10.1109/BigData.2018.8622125
34. Wankhede S, Kshirsagar D. DoS attack detection using machine learning and neural network. In: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. Pune: IEEE (2018). doi: 10.1109/ICCUBEA.2018.8697702
35. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv Preprint.* (2016) arXiv:1608.03983.
36. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Sardinia: JMLR Workshop and Conference Proceedings (2010). p. 249–56.
37. Sumaiya Thaseen I, Saira Banu J, Lavanya K, Rukunuddin Ghalib M, Abhishek K. An integrated intrusion detection system using correlation-based attribute selection and artificial neural network. *Trans Emerg Telecommun Technol.* (2021) 32:e4014. doi: 10.1002/ett.4014
38. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, LA: IEEE (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
39. Ba JL, Kiros JR, Hinton GE. Layer normalization. *arXiv preprint*. (2016) arXiv:1607.06450.
40. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv Preprint*. (2014) arXiv:1412.6980.
41. Patel D, Srinivasan K, Chang CY, Gupta T, Kataria A. Network anomaly detection inside consumer networks—a hybrid approach. *Electronics.* (2020) 9:923. doi: 10.3390/electronics9060923
42. Vincent PD, Sathiyamoorthy E. A novel and efficient key sharing technique for web applications. In: *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE (2013). doi: 10.1109/ICCCNT.2013.6726576
43. Vincent PD. RSA encryption algorithm-A survey on its various forms and its security level. *Int J Pharm Technol.* (2016) 8:12230–40.
44. Bhattacharya S, Maddikunta PK, Kaluri R, Singh S, Gadekallu TR, Alazab M, et al. A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. *Electronics.* (2020) 9:219. doi: 10.3390/electronics9020219
45. Iwendi C, Jalil Z, Javed AR, Reddy T, Kaluri R, Srivastava G, et al. Keysplitwatermark: zero watermarking algorithm for software protection against cyber-attacks. *IEEE Access.* (2020) 8:72650–60. doi: 10.1109/ACCESS.2020.2988160
46. Islam N, Farhin F, Sultana I, Kaiser MS, Rahman MS, Mahmud M, et al. Towards machine learning based intrusion detection in IoT networks. *Comput Mater Contin.* (2021) 69:1801–21. doi: 10.32604/cmc.2021.018466
47. Karatas G, Demir O, Sahingoz OK. Deep learning in intrusion detection systems. In: *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT)*. Ankara: IEEE (2018). p. 113–6. doi: 10.1109/IBIGDELFT.2018.8625278
48. Vincent PM, Sathiyamoorthy E. A secured and time efficient electronic business framework based on public key cryptography. *Int Rev Comp Softw.* (2014) 9:1791–8. doi: 10.15866/irecos.v9i10.3781
49. Iwendi C, Maddikunta PK, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw Pract Exp.* (2021) 51:2558–71. doi: 10.1002/spe.2797
50. Yulianto A, Sukarno P, Suwastika NA. Improving adaboost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset. *J Phys.* (2019) 1192:012018. doi: 10.1088/1742-6596/1192/1/012018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Akshay Kumaar, Samiayya, Vincent, Srinivasan, Chang and Ganesh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fpubh.2021.816226

An Artificial Intelligence Approach Toward Food Spoilage Detection and Analysis

Ekta Sonwani[1], Urvashi Bansal[1], Roobaea Alroobaea[2], Abdullah M. Baqasah[3] and Mustapha Hedabou[4]*

[1] Department of Computer Science and Engineering, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India,
[2] Department Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia,
[3] Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia,
[4] School of Computer Science, Mohammed VI Polytechnic University, Ben Guerir, Morocco

Aiming to increase the shelf life of food, researchers are moving toward new methodologies to maintain the quality of food as food grains are susceptible to spoilage due to precipitation, humidity, temperature, and a variety of other influences. As a result, efficient food spoilage tracking schemes are required to sustain food quality levels. We have designed a prototype to track food quality and to manage storage systems at home. Initially, we have employed a Convolutional Neural Network (CNN) model to detect the type of fruit and veggies. Then the proposed system monitors the gas emission level, humidity level, and temperature of fruits and veggies by using sensors and actuators to check the food spoilage level. This would additionally control the environment and avoid food spoilage wherever possible. Additionally, the food spoilage level is informed to the customer by an alert message sent to their registered mobile numbers based on the freshness and condition of the food. The model employed proved to have an accuracy rate of 95%. Finally, the experiment is successful in increasing the shelf life of some categories of food by 2 days.

Keywords: machine learning for health, smart system, food spoilage detection, food spoilage prevention, sensors, IoMT

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Ebuka Ibeke,
Robert Gordon University, United Kingdom
Michael Edeh,
Ebonyi State University, Nigeria
Mohammad Shah,
University of Bolton, United Kingdom

***Correspondence:**
Urvashi Bansal
urvashi@nitj.ac.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *16 November 2021*
Accepted: *14 December 2021*
Published: *12 January 2022*

Citation:
Sonwani E, Bansal U, Alroobaea R, Baqasah AM and Hedabou M (2022) An Artificial Intelligence Approach Toward Food Spoilage Detection and Analysis. Front. Public Health 9:816226. doi: 10.3389/fpubh.2021.816226

1. INTRODUCTION

Food wastage has been a topic of concern in recent years, and studies are being conducted to identify innovative ways to reduce it. It has been described as a major concern in the long-term sustainability of food production, demand, and food supply chains. Since meals are the essential source of diet for all living beings, the quality and security of meals have always been in high demand. The Internet of Things (IoT) links anything, everywhere, and at any time (1–4) and by incorporating the IoT into the Food Supply Chain (FSC) management, it is possible to improve food shelf life by measuring and monitoring the state of the food and exchanging data to and from customers. Currently, the whole use of IoT technologies in FSC is only in its early stages, with a long way to go (5, 6). To avoid food wastage, food sanitation and safety are of paramount importance. The consistency of meals should be regulated. As a result, wastage can be reduced by installing quality control systems in grocery stores. It will further help in controlling diseases (7). These types of quality control systems keep an eye on the environmental conditions that may maintain the food quality. Previously, atmospheric effects were monitored by procedures such

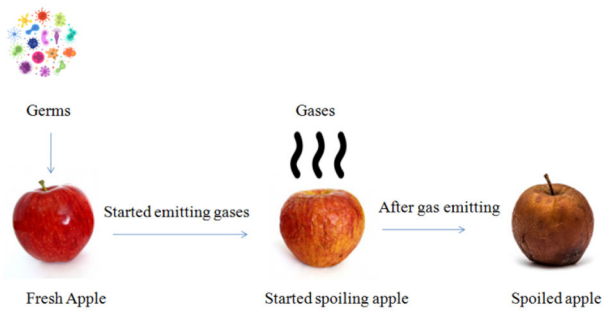


FIGURE 1 | Food spoilage process.

as refrigeration, vacuum storage, and so on. Food pollution can occur during the manufacturing process, although it is most often caused by inefficient food handling activities due to unsuitable environmental conditions during food transportation and storage.

1.1. Food Spoilage Process

Food spoilage occurs when a food product becomes unfit for use by the customer. The factors of such a process are a result of a variety of external factors, including the type of food item, how it is packaged and processed. Every year, one-third of the world's melded are lost which is provided for human consumption due to meal wastage (8, 9). In **Figure 1**, it could be visualized that once the germs start attacking fresh apples, gas emission takes place and after few hours, the apple is spoiled completely.

Bacteria, virus, protozoa, and fungi are factors of food spoilage. These factors can create harmful results for consumers, but we can apply prevention techniques to them to save the life and quality of food. Usually, bacteria can not cause food poisoning and most of the microorganisms that cause food borne illness are odorless and tasteless, otherwise owing to mycotoxins and microbial wastes. Therefore, eating spoiled food is never recommended. Two pathogenic bacteria *Clostridium perfringens* and *Bacillus cereus* can spoil the food.

1.2. Key Component of Food

There is a very important key component i.e., seed. If we want healthy fruits and veggies, that seed must be strong and healthy. The fruits and veggies completely depend on the seeds. There are two types of seeds:

- **Hybrid seeds:** Cross-pollinated plants produce hybrid seeds in cultivation and gardens. Hybrids are selected for their ability to enhance the characteristics of the resulting plants, such as yield, uniformity, color, and disease resistance.
- **Non-hybrid seeds:** These seeds are called heirloom seeds or free pollinated seeds. Non-hybrid seeds originate from naturally pollinated trees. Any of these cultivars have been around for millennia.

The hybrid seeds give hybrid fruits and vegetables and non-hybrid seeds give non-hybrid fruits and veggies. The lifetime of hybrid fruits and veggies is less than the non-hybrid fruits and veggies. But everyone cannot identify which fruit or vegetable is hybrid or not; that is why everybody needs to know the knowledge of their routine work.

Food poisoning can be caused by a variety of factors, including humidity and temperature fluctuations. As a result, it is important to provide a measurement device that can measure humidity and temperature differences during food preparation and transportation (10). Currently, nearly everyone is influenced by the foods they consume daily, not only because of junk food, but also because of canned vegetables and other food items eaten in daily life, which lack consistency because their oxygen, temperature, and moisture content differ from time to time. In the smart house, devices are installed which have the capacity to diagnose wastage of food and then alert caregivers. The study and execution of repetitive measurements, which are aimed at detecting improvements, do not guarantee the nutritional content of the food (11). The details gathered by tracking and monitoring should be reviewed and properly submitted to the administrator for purposes of policy research, pattern forecasting, program assessment, and planning.

The expanded use of smart phones, connected devices, and controllers in food production and other domains has wreaked havoc across the globe. The advantages of IoT technology are rapidly being tapped by the food industry and other sectors (12–14). It accesses *via* broadband (modem) and begins scanning input from the wired up to sensors, the heat, and humidity sensor until properly mounted and turned on. The temperature and moisture controller sensor is a wireless sensor with an active moisture controller sensor. We have employed this idea in the proposed scheme.

Again, the primary function of the proposed monitoring and control systems is to keep track of a certain operation and ensure that it continues as desired. The surveillance process can be accomplished using a variety of different types of sensors. The data collected from the Arduino-based sensors will be compared with the target values (15). If the reading value of the sensor is found to be different from the target values, the control circuit will intervene to affect the allocated operation to keep it at the desired level of quality. As a result, the Smart food spoilage monitoring system focuses primarily on healthy food storage by surveilling and regulating a variety of parameters that affect food products. This control system employs storage units that are equipped with a variety of sensors that read the parameters influencing food quality (16).

1.3. Reasons for Food Spoilage

The recycling bins and the waste made are evidence of the food spoilage circumstances. The big reason for food spoilage is that a larger portion of food waste occurs in office eateries, small and large roadside eateries, group get-together events, and wedding receptions. These food surpluses are not only a widespread indicator of toxicity to the earth's atmosphere, but they also pose a slew of economic problems. According to current calculations, half of all of these foods are lost globally; the global amount of meal wastage is estimated as 1.3 billion tons, and it is projected to rise even more in the upcoming year, which is a concerning problem (17).

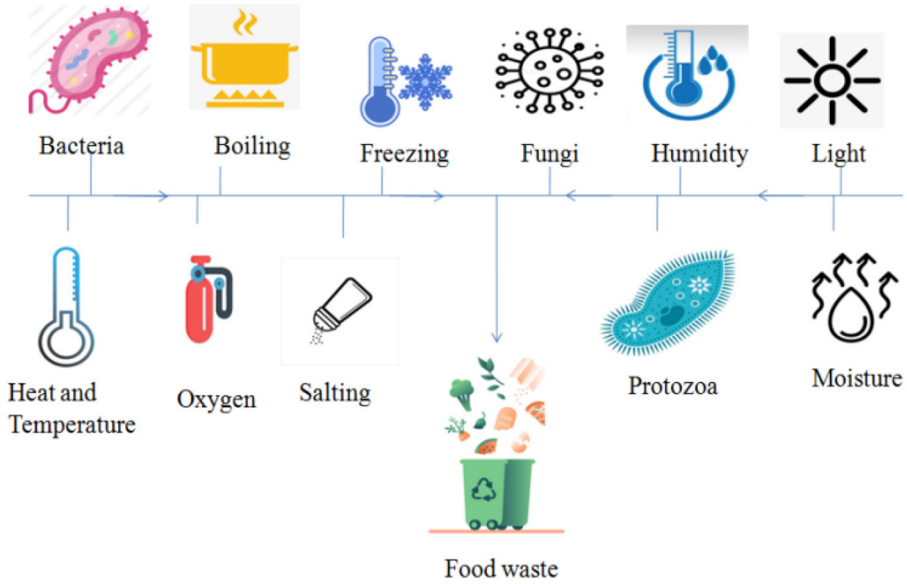


FIGURE 2 | Factors of food spoilage.

Food spoilage can be caused by a variety of human, chemical, and biological causes, including plants, enzymes found in plant food tissues, insects, parasites, and microbes (18). The most common and serious cause of food spoilage is microbial degradation. Food will invariably be polluted by such forms and amounts of microorganisms when microbes are commonly spread in nature.

Figure 2 shows the factors that can destroy fruits and veggies. In terms of heat, humidity, moisture, temperature, and oxygen content they have a threshold value. If the threshold value crosses then food spoiling starts. If we maintain a threshold value then we can vary the life of fruits and vegetables.

1.4. Motivation

The motivation behind Monitoring and analysis of food spoilage using Machine Learning is to keep track and manage food products in order to avoid spoilage caused by climatic and atmospheric changes. Food wastage concerns are focused on healthy eating habits and also concerns with the quality of food as there have been instances where poisonous chemicals have been discovered in fruits and veggies. Monitoring and analysis of food spoilage using Machine Learning saves time and provides accurate and consistent results.

1.5. Our Contribution

While reviewing the articles on food wastage, we have found that there are a few contributions to increasing the shelf life of food. To this end, we have found the research questions for this field. The major and novel contributions of this study are:

- Designing a smart system to keep track of the quality of food using gas, humidity, and temperature sensing mechanism.
- Designing an alert generation system to alert the user about the quality of food and the time spans of its deterioration.
- Development of a system to increase shelf life of food by maintaining the environment according to food.

2. LITERATURE REVIEW

Continuous sensing in a smart home is a unique sensor modality that has the ability to distinguish diverse everyday activities and indicate potentially dangerous situations for the person who lives there. One such situation is food rotting, which is the subject of this study (19). The authors have evaluated the smell signatures of two typical meals (milk and yogurt) that were kept at 25°C for a week using a metal-oxide sensor (MOS) based electronic nose. The highest absolute sensor responses were used to create feature vectors, and their components followed a smooth trajectory as the data aged. The two chemicals followed different trajectories during spoiling, according to principal component analysis (PCA).

Rajesh Megalingam et al. introduce a unique method for detecting food deterioration by combining picture classification with machine learning techniques and artificial intelligence (20). They have used AI, deep CNN networks, computer vision, and ML techniques such as the k clusters method for color classifications in pictures and its HSV values for spoiling detection to identify food rotting. The anaconda prompt is used to complete this project on the jupyter notebook platform. Also, Iwendi et al. employed an Artificial Intelligence approach for detecting and analyzing security levels on IoT using a network classifier. The study shows an impressive accuracy rate for the proposed application. In this view, we can say that AI is useful in almost every field nowadays.

Green et al. presented an electronic nose (e-nose) that consists of four gas sensors made of functionalized single-walled carbon nanotubes (f-SWNTs) and polymer nanocomposites, with the goal of becoming a simple monitoring device for microbiological spoilage and pollutants in canned food (21). The gas sensing signals were utilized as early indications of deterioration to assist in averting negative health impacts. In a static environment, the gas sensor array was tested to see how it responded to different volatile organic compounds (VOCs). Because they have an excellent sensing response to ammonia, which is one of the gases generated by microorganisms. These sensors are suitable for detecting microbial deterioration of canned food. The sample, which was canned tuna in mineral water, was opened and kept at room temperature (25°C). The odor linked with the deterioration of canned tuna was then monitored for 10 days using e-nose. PCA was also used to visualize the discrimination of microbial canned tuna spoilage status and analyze the smellprint of a specific level of ammonia contamination. PCA was found to be able to track changes in canned tuna deterioration, demonstrating its potential to be used for quality assurance of canned food in the daily life of a smart home resident.

When food is reversed in an inadvertent situation, such as an insufficient temperature, foodstuff rots as a result of the rapid breeding of food spoilage bacteria in warm and wet conditions where bacteria can easily be generated (22). The consequences can be even worse if humans eat those degenerative foodstuffs, as they may develop bromatoxism. The goal of this study is to design and simulate a wearable RFID patch for food spoilage monitoring with smart packaging that can be recognized and read temperature information by a device that supports near NFC technology *via* an attached circular antenna.

Everyone has a distinct opinion on whether or not a meal is rotten, which can lead to incorrect conclusions about the item's state (23). Foodborne sickness might result from a misinterpretation of the food condition. Due to this forecast ambiguity, there is still a desire for an electronic nose system to categorize if a meal is rotten. The goal of this study is to use an electronic nose to identify food deterioration in tomato-based Filipino cuisines. This study seeks to create a device with an array of sensors that can detect gases released by rotten tomato-based Filipino cuisines and to use an Artificial Neural Network as an algorithm to classify the sensor's data readings. The e-nose system's hardware includes a Gizduino 1281, a Raspberry Pi 3 Model B, a 20x4 LCD screen, seven MQ gas sensors, and one temperature/humidity sensor. The Artificial Neural Network data (24, 25) is trained using Stochastic Gradient Descent and the Back Propagation technique. The sensor chamber is positioned underneath the tomato-based Filipino food for simple detection of the gas sensors generated by the cuisine. The research might be beneficial in determining whether or not food has gone bad. This electronic nose device may detect the degree of rotting in a certain tomato-based Filipino dish. The researchers assigned rotting levels ranging from 0 to 12. Starting at 7:00 a.m. on Day 1, these levels correlate to the period when the meal is monitored every 4 h. As the trial progressed, the food deteriorated often between levels 5 and 6. The electronic nose system has a 3.85% mistake rate, according to the confusion matrix.

Benjamin et al. described a flexible UHF RFID sensor for detecting food quality (26). This sensor is based on inter-digital capacity found in an RFID antenna, onto which a coating of vegetal biopolymer has been placed. Electromagnetic coupling between the capacity and the biopolymer is, therefore, employed to vary the adaption coefficient between the chip and the antenna of the RFID tag based on food deterioration. Experimental measurements of an RFID-sensor exposed to a genuine food gas environment in the process of deterioration.

Food losses after harvest can range from 10 to 30%, owing to a variety of fungal and other microbes that degrade the food during transportation and storage (27). In addition to these direct losses due to spoiling, several bacteria generate natural poisons that make foods unsuitable for human ingestion. Other fruits and vegetables require waxes and water-loss barriers on the surface to retain quality and ideal sensory characteristics between farm and market. This study describes a study and development of an effective electrostatic spray application technology as well as a processing-line prototype designed particularly for food postharvest protection. The study provides experimental findings of comprehensive assessments of electrostatically applied protective sprays onto bananas for international transport, where both microbiological and mass-transfer data show usually 2.1–3.4-fold deposition benefits for food protection.

Microbial deterioration in milk is a major source of concern for those who want to eat healthily (28). The goal of this research is to use an electronic nose system (e-nose) with a variety of gas sensors as an initial tool for monitoring food deterioration in order to preserve food safety and human health. Nanoparticles gas sensors made of various polymers and f-SWCNTs are used in this device. When these gas sensors are exposed to the headspace of the milk, they respond to the VOCs present in the milk. Pasteurized milk was stored at 4 and 25°C for 9 days to monitor the development of spoiling. The odor pattern was examined using PCA to assess the capabilities of an electronic nose for freshness and milk spoiling detection. Monitoring the change in detecting responses revealed each sensor's capacity to identify the strength of the odor level. The findings of the discrimination demonstrate that the odor levels of samples maintained at 4°C do not change from the first to the ninth day, but the odor levels of samples held at ambient temperature increase with storage duration.

Food planning and quality assessments are becoming increasingly essential aspects of modern society (29). We may effectively analyze crop quality utilizing various learning approaches thanks to the scientific vision and agricultural advancements, which certainly replace inconsistent, unpredictable, and time-consuming human labor. The use of fuzzy logic to crop quality has effectively demonstrated its worth and established itself as a viable alternative to automation, indicating that it has a bright future ahead of it. This article examines some of the most important research findings on this subject.

Food safety is currently a key problem for the food business (30). Food businesses must develop quality monitoring systems in order to discover food quality issues early in the production process. Data manipulation and centralized storage, on the other hand, have become obstacles to dependability in a typical quality monitoring system. Furthermore, traditional quality monitoring techniques are typically ineffective due to a lack of effective automation. Fortunately, block-chain is a promising, tamper-proof, and decentralized technology. Furthermore, smart contracts, which are self-executing and self-verifying programs on the blockchain network, may execute transactions between mutually untrustworthy parties. This article offers an intelligent quality monitoring system for fruit juice manufacturing that combines smart contracts and quality rating algorithms. This system offers a high level of automation as well as a high level of dependability. Response surface models are created based on preproduction data in this system, and the best production condition for each step is determined. Smart contracts are used to capture production data on a blockchain during the actual production process.

Food is extremely essential in our day-to-day lives. The quality of food is deteriorating day by day as globalization progresses. To keep the food fresh, various food processing is done most of the time. Various preservatives or substances are added to the meal to give it a fresh or appealing appearance. The majority of food is now preserved using chemicals, resulting in food contamination. As a result of the pollution, numerous illnesses develop, prompting consumers to seek out healthier foods. Organic food is desired by the public in order to maintain a healthy lifestyle. So, in order to prevent difficulties linked with food that cannot be interpreted by humans, we need a gadget that can assist in determining the quality of food. There is a need for such a device that instructs us on how to eat sanitary meals. As a result, in order to meet this customer need, we created a gadget that determines if the food quality is good or terrible. This article examines the application of several sensors in the food business. Sensors such as pH sensors, gas sensors, and temperature sensors aid in determining the state of food. This system has a strong presence in restaurants, homes, and small businesses (31).

3. EXISTING METHODS AND SOLUTIONS

The goal of the smart meal tracking device is to track and regulate food products to avoid harm caused by atmospheric or climate variability. Food waste may also be caused by insufficient storage of foods. Through tracking and regulating different criteria concerning foodstuffs, the smart food monitoring system focuses on healthy food storage. This device employs storage units that have been implanted with a variety of electronic sensors that can interpret the parameters that affect foodstuffs.

3.1. Techniques for Food Preservation

Food preservation's main goal is to save food from spoiling so it can be eaten. Gardeners also grow too much food at once, much more than can be consumed until it spoils. Food preservation also allows you to enjoy a wide range of foods all year long. The four techniques for food preservation are shown in **Figure 3** (32).

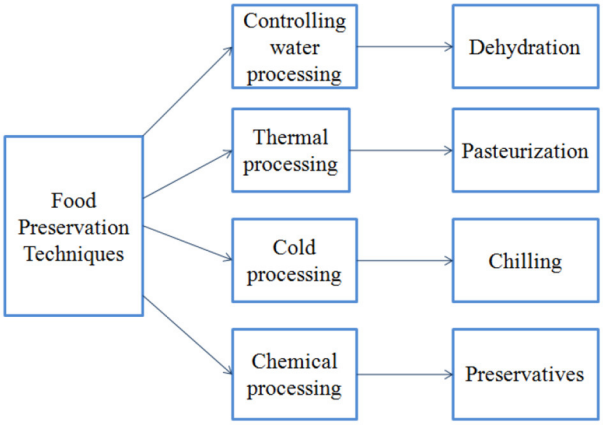


FIGURE 3 | Food preservation techniques.

3.2. Alternatives of Artificial Preservatives

Various natural preservatives derived from plants, insects, fungi, and alloy have antioxidant, antimicrobial, and anti-enzymatic effects in studies. Neem, basil, rosemary, and clove take out are encouraging alternatives for their synthetic equivalent (33).

Studies show that manufacturing companies are using chemical antioxidants which is the cause of hyperactivity in previously non-hyperactive people. Natural compounds derived from plants, organisms, or alloy may be useful substitutes. They may even be found as preservatives in meats, cosmetics, and pharmaceuticals, where they could be used as flavoring, binding, disintegrating, gelling, thickening, or suspension agents, even as automobiles.

Researchers also used extracts from natural sources such as green tea, grape seed, and nisin, a bacteriocin known as a nutritious food preservative, to make chicken and turkey hot dogs in a series of trials (34). **Table 1** represents the information regarding various food preservatives used for different kinds of fruits and vegetables. **Table 2** represents the diseases caused using preservatives.

3.3. Importance of Food Preservation

- Bacterial development and other forms of spoilage are inhibited by preservation processes, ensuring that the food is healthy enough to consume.
- Pickling competes with freezing, canning, and drying as a way to save produce from spoiling. Antioxidants, amino acids, and beneficial bacteria are commonly found in fermented foods.
- Food preservation delays the breakdown of rancid-causing fats while also inhibiting the development of microorganisms (such as yeasts) and other microorganisms.
- Feed storage expands the stock of food.
- Feed storage helps to reduce food waste. Excess foods that would have been discarded otherwise are processed and stored, adding to available stocks and reducing food waste (35).

TABLE 1 | Different types of food containing various kinds of preservatives.

Preservatives	Food
Sorbic acid	Syrups, sweets, dairy products, fruit products, fermented products, beverages
Tert butyl hydroquinone (TBHQ)	Fats, oils, snack foods
Tocopherols (vitamin E)	Oils
Ascorbic acid (vitamin C)	Fruit and acidic products
Butylated hydroxyanisole (BHA) and Butylated hydroxy -toluene (BHT)	Fats and oils, bakery products, cereals
Sodium sorbate	Mayonnaise, processed meats, dairy products, fermented products
Sodium and calcium propionate and Potassium propionate and propionic acid	Breads and other baked goods
Benzoic acid and sodium benzoate	Fruit products, margarine, and acidic foods
Calcium lactate	Olives, frozen desserts, jams, jellies, and dairy products
Calcium sorbate	Mayonnaise, dairy products, syrups, and margarine
Ethylene diamine tetra acetic acid (EDTA)	Dressings, canned veggies, and margarine
Methylparaben	Relishes, dressings, and beverages
Propylparaben	Cake, pastries, beverages, and relishes
Sodium nitrate and nitrite	Cured meats, fish, and poultry

- Dietary shortages can be reduced by storing food. Preserved foods help to bring diversity to the diet. In certain Middle East nations, there is a shortage of plantations due to arid environmental conditions. Importing new and dried fruits and veggies makes up for this shortcoming.

3.4. Reasons for Food Preservation

- The product could not be stored for a longer time as it could be spoiled.
- Using a new product that is in season and produced locally, remaking preserving jars, cutting food waste, and reducing food miles are all eco-friendly choices.
- Buy in bulk and seasonally to save costs.
- Artificial preservatives, BPA, and other artificial additives may not be used in home-prepared food.
- Convenience - in minutes, it is ready to eat.
- Both the canning process and sharing with families, colleagues, and neighbors need community participation. Teach the next generation how to do things.
- Have plans for unexpected events and/or injury.
- Personal fulfillment, we enjoy coming up with different flavor blends that you would not find in your grocery supermarket, as well as demonstrating how to use packaged foods in recipes.
- Learn new skills, such as how to use all of the fruit and how to cut down on food waste. Consider the following scenario: Apples - cut into slices (or use as pie filling) and squeeze the juice from the peels and cores.

4. ARCHITECTURE

The proposed solution includes new device architecture with few configuration modifications as per the requirement. The new device architecture includes a humidifier to maintain the humidity of the device and a cooling module that maintains the temperature for the food items. The prototype of the device is represented in **Figure 4**. Various electronic instruments have been used for monitoring. Furthermore, the registered values are used for monitoring purposes (36). The information gathered by various sensors can be correlated to the target values. The speed control kicks in and interprets the provided procedure to keep it running in the correct direction if the sensor readings do not match the key parameters. This theory can be used to create a method that can protect raw foods. This architecture consists of the following components:

- **Controller:** Arduino is used as a microcontroller.
- **Gas detection sensor:** A gas detector is a device that measures the number of gases in a given environment, and is often used as part of a protection system. Operators in the field where the leak is happening will be alerted by a gas detector (37, 38).
- **Humidifier:** A humidifier is a system that raises humidity (moisture) in a single room or a whole house. It is essentially an electrical appliance. Point-of-use humidifiers are widely used in the home to humidify a single room, and humidifiers are used in medical ventilators for greater patient convenience.
- **Heat sensor:** The key function of a heat sensor is to detect the heat that is present inside it. When the temperature around the heat sensor rises above its fixed point, it detects the heat and alerts us, allowing us to protect the devices from injury (39).
- **Humidity sensor:** A humidity sensor is a sensor that senses humidity and converts the information into an electrical current. Humidity sensors are available in a range of shapes and sizes, as well as with different features (40). The Humidity sensor first senses any humidity or moisture in the fruits or vegetables and it sends some value to the microcontroller if any moisture content is detected then an alert message sends to our mobile phone if the moisture content is not detected then again taking values from the microcontroller.
- **Cooling module (TEC1-12715-Thermoelectric Cooler 15A Peltier Module):** The heater, condenser, and fan device are all included in this element. A heater and a condenser are included in this module.
- **Light sensor:** Light sensors are electronic devices that monitor the intensity of natural or artificial light. Light energy is converted into an electrical signal by these devices. Light sensors are used in a variety of manufacturing application areas.

In **Figure 5**, the Gas detection sensor senses any gas or air coming from the fruits or vegetables, and it sends some value to the microcontroller if any gas content is detected then it alerts the user. If gas content is not detected then again values are taken from the microcontroller.

In **Figure 6**, the Humidity sensor first senses any humidity or moisture in the fruits or vegetables, and it sends some value to the microcontroller. If any moisture content is detected then it alerts

TABLE 2 | Dangerous food preservatives cause various diseases.

Preservatives	Cancer possibility (Yes/No)	Asthma possibility (Yes/No)	Hypersensitivity possibility (Yes/No)
Calcium/Potassium/Sodium propionate and propionic acid	No	Yes	Yes
Sodium and potassium nitrate	Yes	No	Yes
Sodium nitrite	Yes	Yes	Yes
Butylated hydroxyanisole (BHA)	Yes	Yes	Yes
Butylated hydroxytoluene (BHT)	Yes	Yes	Yes
Tert butyl hydroquinonesynthesiz-ed (TBHQ)	No	Yes	Yes
Sodium benzoate	Yes	Yes	Yes
Potassium and calcium sorbate and Sorbic acid	No	Yes	Yes
Benzoic acid	No	Yes	Yes
Propylparaben	No	Yes	No
Sulfur dioxide	No	Yes	Yes
Potassium bisulfite	No	Yes	Yes
Hexamethylen-etetramine	Yes	No	No
Sodium metabisulphite	No	Yes	No

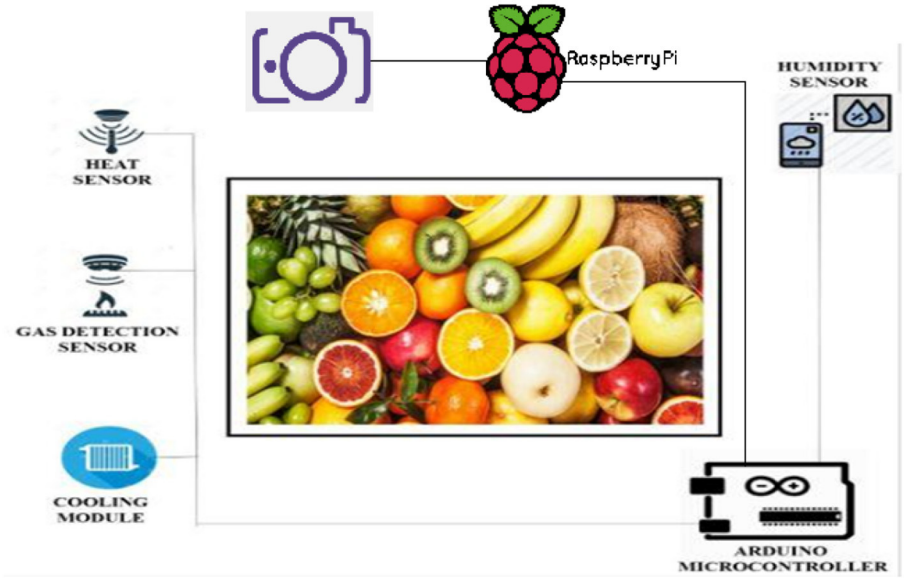


FIGURE 4 | The architecture of Monitoring and analysis of food spoilage using Machine Learning.

the user and if the moisture content is not detected then again values are read by the microcontroller. As shown in **Figure 7**, the heat sensor and cooling module senses any heat in the fruits or vegetables and it sends some value to the microcontroller. The complete process is given in **Algorithm 1**.

5. WORKING PRINCIPLE

We have different types of sensors that play a major role in the detection and monitoring of food spoilage. It includes the camera sensor, humidity sensor, gas sensor, and heat sensor. The camera sensor captures the image of fruit or vegetable. The humidity sensor senses the humidity of the environment. If it is below the threshold value then the humidifier increases the humidity up to the threshold value. The temperature sensor monitors the temperature for the predefined threshold value which is controlled by Arduino. Arduino is directly connected to the Raspberry pi, which acts as a mini computer having own processing unit and memory. Whenever temperature reaches above the value of the threshold, the cooling module gets turned on. The gas sensor detects early spoilage *via* detecting a little amount of gas emission of the food items. It sends the value to the Arduino, after receiving the values, Arduino pushes the values to the cloud. The web application notifies the user *via* a buzzer or voice-activated commands or through the display messages.

6. SMART FOOD LIFE PREDICTION

The food life cycle has different stages which include the following stages: Fresh Food; Food spoilage just started; Spoilage Food. For prediction, we have trained 50 types of fruits and vegetables over different three classes. We proposed a CNN for object detection and prediction model. This is trained over three different classes.

Object detection is a difficult task in the realm of computer vision. Machine learning models when trained on a large enough number of photos can be used to detect and distinguish objects. We will identify the fruit or vegetable by using the Convolutional Neural Network (CNN) to calculate the threshold values for different parameters e.g., gas, temperature, and heat as these values are different for a different type of food. The networks will be trained in a supervised way, with photos of the fruits serving as input and tags serving as a result. The CNN model will be able to correctly anticipate the fruit label after successful training.

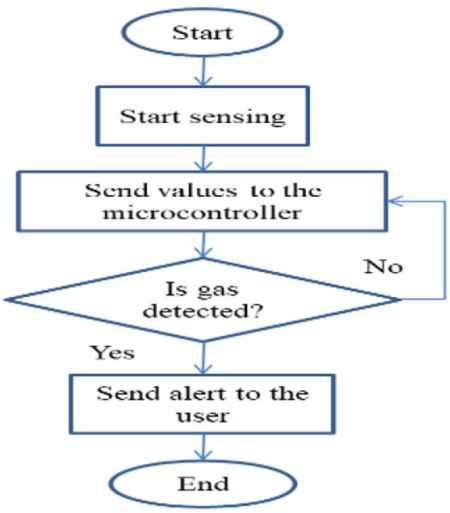


FIGURE 5 | Gas detection sensor.

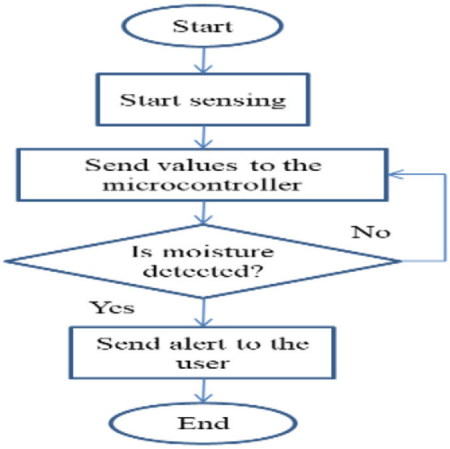


FIGURE 6 | Humidity sensor.

Algorithm 1: Process (object).

1: Turn on the device
2: Capture the image of fruit or vegetable
3: Turn on the Cooling module
4: Turn on Humidifier
5: Store the optimal values of parameter according to captured object
6: Read the values of sensor for monitoring process of fruits or vegetables
7: **if** Gas content is detected **then**
8: go to step 20
9: **else**
10: go to step 21
11: **end if**
12: **if** Moisture content is detected AND moisture > maximum optimal humidity of object AND Humidifier is off **then**
13: Turn on the humidifier
14: **else if** heat content is detected AND temperature > maximum optimal storage temperature of object AND cooling module is off **then**
15: Turn on the cooling module
16: **else if** life of object > average shelf life of object **then**
17: go to step 20
18: **else**
19: go to step 21
20: **end if**
21: Send alert to the user
22: Capture an image of fruit or vegetable
23: go to step 7

We have utilized the deep learning approach to train our model on datasets of fruits and vegetables. We used the Fruits360 dataset, which is openly available and may be downloaded from GitHub or found on Kaggle (41). It has around 90,380 pictures in various categories. This dataset contains high-resolution pictures, which are required for a successful classifier.

In our model, there are a total of 11 layers. The output layer is SoftMax, and there are four convolutional layers, four max pooling layers, and two fully connected layers. The input layer is a convolutional layer with a 16, 5 x 5 x 4 filter and 2 x 2 max-pooling with stride = 2. Convolutional layer 4 has a parameter of (5 x 5 x 64) and a dimension of 128. The model has been trained on 50 different types of fruits and vegetables, and it is also capable of identifying multiclass images.

7. RESULTS AND DISCUSSION

Information gathering is an important and time-consuming task. Regardless of the field of study, precise data collection

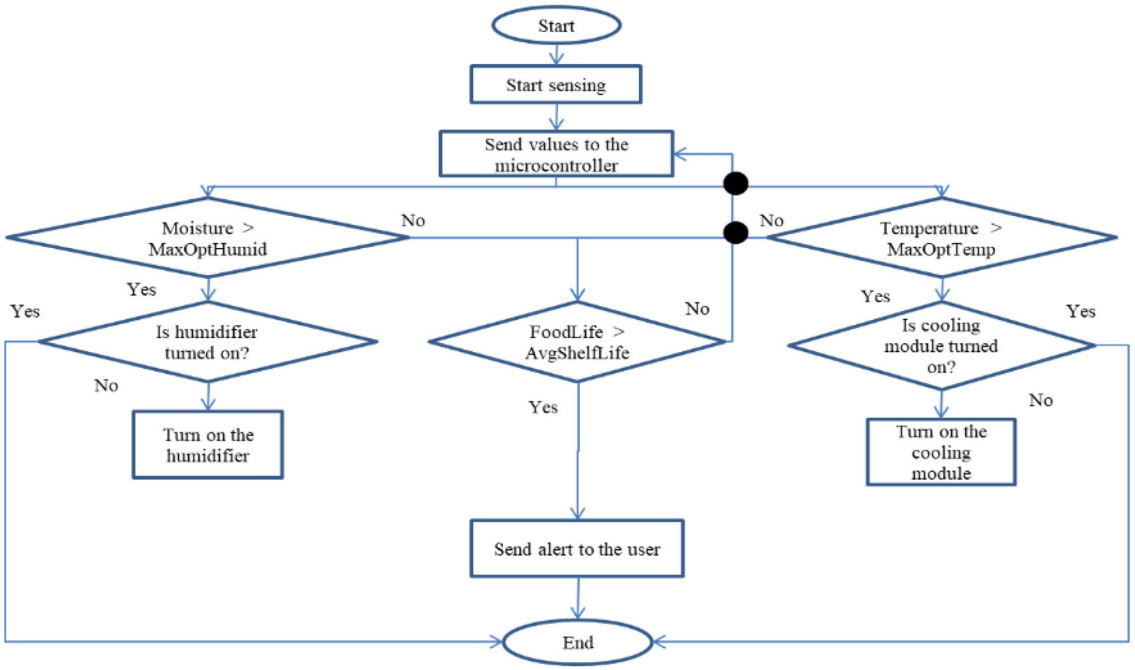


FIGURE 7 | Heat sensor and cooling module.

TABLE 3 | Description of the fruits and veggies dataset.

Columns	Description	Value	Type
Names of fruits and vegetables	Different types of fruits and vegetables	NA	String
Minimum optimal storage temperature	Minimum temperature in which fruit or vegetable remain fresh	Multiple minimum optimal temperature values	Numeric
Maximum optimal storage temperature	Maximum temperature in which fruit or vegetable remain fresh	Multiple maximum optimal temperature values	Numeric
Freezing point	This cooling point in which fruit or vegetable remain fresh	Multiple freezing point values	Numeric
Minimum optimal humidity	Minimum humidity in which fruit or vegetable remain fresh	Multiple minimum optimal humidity values	Numeric
Maximum optimal humidity	Maximum humidity in which fruit or vegetable remain fresh	Multiple maximum optimal humidity values	Numeric
Minimum approximate storage life	At least number of days in which fruit or vegetable remain fresh	Multiple minimum approximate storage life values	Numeric
Maximum approximate storage life	At most number of days in which fruit or vegetable remain fresh	Multiple maximum approximate storage life values	Numeric
Average shelf life	Average of minimum (start spoiling) spoilage time and maximum (after spoiled) spoilage time	Multiple average shelf life values	Numeric

is crucial to maintaining cohesion. Food spoilage is a very concerning topic nowadays because this time is needed to have a strong immune system, and there is no existing solution that helps to prevent spoilage to some extent. The Fruit360 is the data set that was used in this study. We have to gather a dataset from the Kaggle dataset website. The dataset contains images of different fruits and vegetables. We have used this dataset for CNN model training and object detection.

Fruits and Veggies—We have created a dataset after careful research. The dataset contains 50 types of fruits and vegetables with their optimal storage temperature range, optimal humidity range, approximate storage life range, and freezing point are included in this collection (42). Dataset is used for spoilage alert

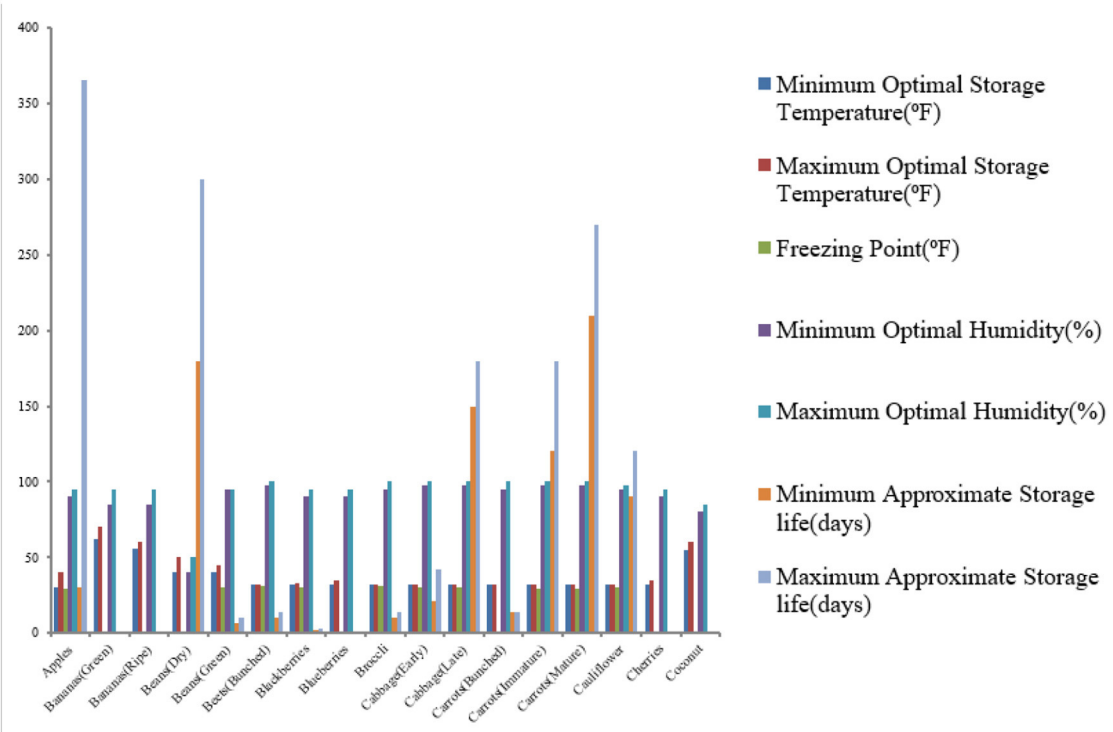


FIGURE 8 | Comparison among different fruits and vegetables with respect to their various attributes.

generation messages in mobile. **Table 3** describes the dataset attribute with its value, type, and description. All the value of each column represents numeric and floating point values except the first column.

Figure 8 represents the relation between attributes like minimum optimal storage temperature, maximum optimal storage temperature, Freezing Point, minimum optimal humidity, maximum optimal humidity, minimum approximate storage life, and maximum approximate storage life of 17 fruits and vegetables. We have selected 15 fruits or vegetables out of 50 fruits or vegetables for our experimental purpose. This figure represents the comparison among standard parameter settings of different food items. Then we put the food item in the proposed device, which further analyzes the item and sets the parameter values accordingly. Finally, the proposed approach is used to monitor the item and maintain the environment accordingly. In the end, we monitored the final shelf life of the food item. In this experiment, we try to extend the maximum approximate storage life of fruits and vegetables, which is represented in **Table 4**.

The average shelf life of any fruit or vegetable is a time period in which fruits or vegetables are healthy for eating. After this period of time, it is harmful for health to consume this fruit or vegetable as spoilage may have taken place previously before the average shelf life. Every fruits or vegetables have a different minimum optimal storage temperature below which they started spoiling. So, careful minimum optimal storage temperature must be maintained.

The performance graph for our model is plotted for both the training and validation data. We show the training data loss vs. validation data loss graph, as well as the training data accuracy and validation data accuracy of the CNN model for fruits and vegetables, for object detection. The performance graph plot shows the improvement in accuracy as the number of epochs increases, as well as whether or not our model has been correctly trained. The CNN model's training data loss vs. validation data loss, as well as training data accuracy and validation data accuracy, are shown on the graph shown in **Figures 9**, **10**. There are three different fit situations for the model: over fitting, under fitting, or perfect fit.

- **Over fitting:** When the Training loss is considerably less than the Validation loss, this scenario occurs.
- **Under fitting:** This situation occurs when the Training loss exceeds the Validation loss, i.e., Training loss > Validation loss.
- **Perfect fit:** This circumstance occurs when Training and Validation losses are approximately equal or converging over time, indicating that we are doing things correctly.

As seen in our graph of **Figure 9**, it is virtually converging to indicate that our model is free of over fitting and under fitting. The plot of training and validation accuracy after each epoch is shown in **Figure 10**, our model is virtually ideal in this case since the validation and training accuracy are nearly identical. So, our model has a 95% accuracy rate, which is excellent enough to recognize the kind of vegetable and fruit from a picture. It also

TABLE 4 | Experimental analysis of fruits and vegetables.

Name of fruits or vegetables	Minimum temperature (°F)	Maximum temperature (°F)	Average shelf life (days)	Maximum approximate storage life (days)	After experimental analysis (days)
Broccoli	32	32	11	14	16
Cabbage (Early)	32	32	41	42	44
Carrots (Immature)	32	32	35	180	181
Cauliflower	32	32	14	120	122
Cherries	30	31	6	14	15
Grapes	31	32	6	56	55
Kohlrabi	32	32	7	90	91
Gooseberries	31	32	3	28	29
Leeks	32	32	11	90	91
Parsley	32	32	6	90	91
Plums	31	32	4	35	36
Eggplant	46	54	2	7	9
Blackberries	32	33	6	3	4
Corn (Sweet)	32	32	7	8	9
Cucumbers	50	55	11	14	15

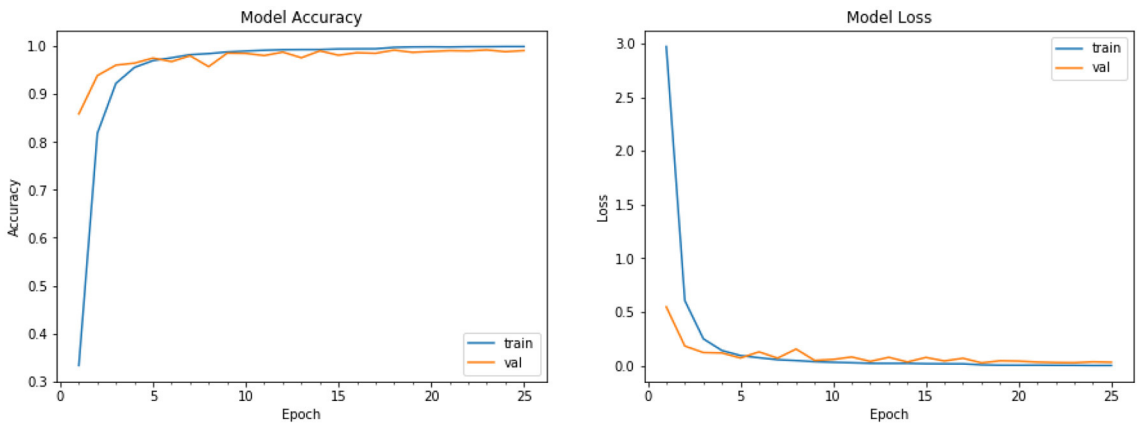


FIGURE 9 | Performance graph of Training Loss vs. Validation Loss.

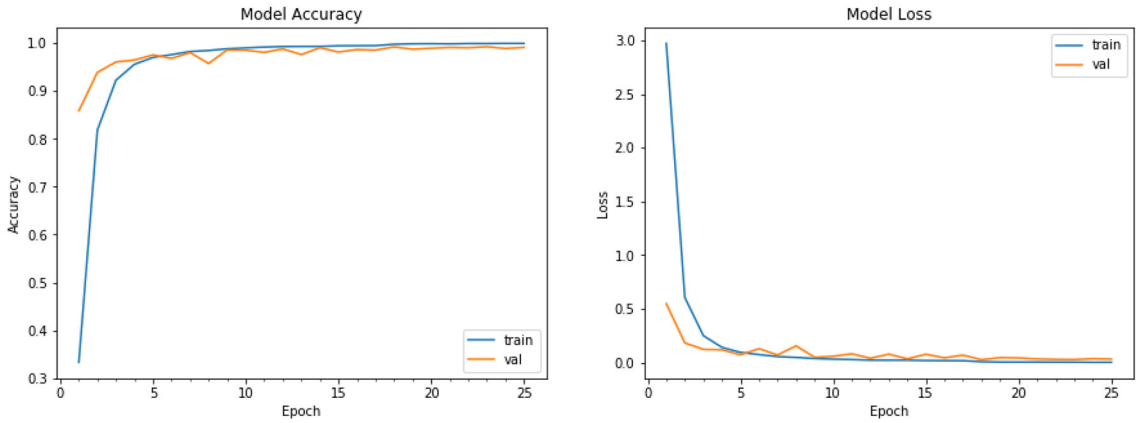


FIGURE 10 | Performance graph of training accuracy vs. validation accuracy.

works well for single items, such as single fruits or vegetables. It also has the ability to forecast the class of numerous objects. For 50 different sorts of categories, the approach works perfectly. It provides extremely precise results for the photos that are fed into the model.

8. CONCLUSION AND FUTURE SCOPE

This study presents a novel technique for Monitoring and analysis of food spoilage using a sensor bases system. The device proposed in this study is able to preserve food for more days. Additionally, food items can be prevented from getting spoiled by increasing their lifespan. It monitors the quality of food items and keeps notifying the user with voice-activated commands or *via* display, and it also generates alerts to the user with the predicted remaining time of the food spoilage. The proposed device shows an accuracy of 95%.

The proposed smart device can be improved by applying image processing and machine learning algorithms to detect early spoilage. This can be utilized in refrigeration systems for detecting food items, spoilage and monitoring for prevention of food spoilage. The device can be incorporated into food transportation containers which would allow tracking and detecting the spoilage if any during transportation. The device could also be tested for different variety of foods as well.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://docs.google.com/spreadsheets/d/1Cz8N7z6tJ8FouYnbbn5shy2D71Ln5P0W/edit?usp=sharing&ouid=105439467837665352046&rtpof=true&sd=true.

AUTHOR CONTRIBUTIONS

UB and ES carried out the experiment. ES wrote the manuscript with support from UB. ES fabricated the device with the help from RA, AB, and MH. UB supervised the project. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors are grateful to the Taif University Researchers Supporting Project (Number TURSP-2020/36), Taif University, Taif, Saudi Arabia and Dr B. R. Ambedkar National Institute of Technology Jalandhar for their support to carry out this work.

REFERENCES

1. Dhanamjayulu C, Nizhal U, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Proc*. (2021). 1–12. doi: 10.1049/ipr2.12222
2. Ahmad P, Jin H, Alroobaea R, Qamar S, Zheng R, Alnajjar F, et al. MH UNet: a multi-scale hierarchical based architecture for medical image segmentation. *IEEE Access*. (2021) 9:148384–408. doi: 10.1109/ACCESS.2021.3122543
3. Arul R, Alroobaea R, Mechti S, Rubaiee S, Andejany M, Tariq U, et al. Intelligent data analytics in energy optimization for the internet of underwater things. *Soft Comput*. (2021) 25:12507–19. doi: 10.1007/s00500-021-06002-x
4. Anajemba JH, Tang Y, Iwendi C, Ohwoekevwo A, Srivastava G, Jo O. Realizing efficient security and privacy in IoT networks. *Sensors*. (2020) 20:2609. doi: 10.3390/s20092609
5. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*. (2020) 9:274. doi: 10.3390/electronics9020274
6. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intell Humanized Comput*. (2020) 1–14. doi: 10.1007/s12652-020-01963-7
7. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in pakistan. *Comput Mater Continua*. (2021). 69:1253–69. doi: 10.32604/cmc.2021.015655
8. Jedermann R, Nicometo M, Uysal I, Lang W. *Reducing Food Losses by Intelligent Food Logistics*. London: The Royal Society Publishing (2014).
9. Brody AL, Bugusu B, Han JH, Sand CK, McHugh TH. Innovative food packaging solutions. *J Food Sci*. (2008) 73:107–116. doi: 10.1111/j.1750-3841.2008.00933.x
10. Pang Z, Chen J, Zhang Z, Chen Q, Zheng L. Global fresh food tracking service enabled by wide area wireless sensor network. In: *2010 IEEE Sensors Applications Symposium (SAS)*. Limerick: IEEE (2010). p. 6–9.
11. Drago E, Campardelli R, Pettinato M, Perego P. Innovations in smart packaging concepts for food: an extensive review. *Foods*. (2020) 9:1628. doi: 10.3390/foods9111628
12. Bourouis S, Alroobaea R, Rubaiee S, Andejany M, Almansour FM, Bouguila N. Markov chain monte carlo-based bayesian inference for learning finite and infinite inverted beta-liouville mixture models. *IEEE Access*. (2021) 9:71170–83. doi: 10.1109/ACCESS.2021.3078670
13. Iwendi C, Khan S, Anajemba JH, Bashir AK, Noor F. Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access*. (2020) 8:28462–74. doi: 10.1109/ACCESS.2020.2968537
14. Hebbar N. Freshness of food detection using IoT and machine learning. In: *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. Vellore: IEEE (2020). p. 1–3.
15. Kumar A. An arduino sensor-based approach for detecting the food spoilage. *Int J Eng Appl Sci Technol*. (2020) 5:596–9.
16. Chander BR, Kumari G. Food quality monitoring system by using arduino. *J Eng Sci*. (2020). 11:625–30.
17. Ishangulyyev R, Kim S, Lee SH. Understanding food loss and waste–Why are we losing and wasting food? *Foods*. (2019) 8:297. doi: 10.3390/foods8080297
18. Yousefi H, Ali MM, Su HM, Filipe CD, Didar TF. Sentinel wraps: real-time monitoring of food contamination by printing DNAzyme probes on food packaging. *ACS nano*. (2018) 12:3287–3294. doi: 10.1021/acsnano.7b08010

19. Green GC, Chan AD, Goubran RA. Tracking food spoilage in the smart home using odour monitoring. In: *2011 IEEE International Symposium on Medical Measurements and Applications*. Bari: IEEE (2011). p. 284–7.
20. Megalingam RK, Sree GS, Reddy GM, Krishna IRS, Suriya L. food spoilage detection using convolutional neural networks and k means clustering. In: *2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE)*. Noida: IEEE (2019). p. 488–93.
21. Green GC, Chan AD, Goubran RA. Monitoring of food spoilage with electronic nose: potential applications for smart homes. In: *20093rd International Conference on Pervasive Computing Technologies for Healthcare*. London: IEEE (2009). p. 1–7.
22. Yuan M, Ghannam R, Karadimas P, Heidari H. Flexible RFID patch for food spoilage monitoring. In: *2018 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia)*. Chengdu: IEEE (2018). p. 68–71.
23. Caya MVC, Cruz FRG, Blas PJR, Cagalingan MM, Malbas RGL, Chung WY. Determining spoilage level against time and temperature of tomato-based Filipino cuisines using electronic nose. In: *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*. Manila: IEEE (2017). p. 1–5.
24. Ponnan S, Saravanan AK, Iwendi C, Ibeke E, Srivastava G. An artificial intelligence based quorum system for the improvement of the lifespan of sensor networks. *IEEE Sens J*. (2021) 21:17373–85. doi: 10.1109/JSEN.2021.3080217
25. Jorge-Martinez D, Butt SA, Onyema EM, Chakraborty C, Shaheen Q, De-La-Hoz-Franco E, et al. Artificial intelligence-based Kubernetes container for scheduling nodes of energy composition. *Int J Syst Assurance Eng Manag*. (2021). 1–9. doi: 10.1007/s13198-021-01195-8
26. Saggin B, Belaizi Y, Vena A, Sorli B, Guillard V, Dedieu I. A Flexible Biopolymer based UHF RFID-Sensor for food quality monitoring. In: *2019 IEEE International Conference on RFID Technology and Applications (RFID-TA)*. Pisa: IEEE (2019). p. 484–7.
27. Law SE, Cooper SC. Air-assisted electrostatic sprays for postharvest control of fruit and vegetable spoilage microorganisms. *IEEE Trans Ind Appl*. (2001) 37:1597–602. doi: 10.1109/28.968166
28. Phukkaphan N, Eamsa-ard T, Chairanit C, Kerdcharoen T. The application of gas sensor array based electronic nose for milk spoilage detection. In: *2021 7th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*. Pattaya: IEEE (2021). p. 273–6.
29. Goel P, Goel S, Bhatia S. Food quality assessment using fuzzy logic. In: *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. New Delhi: IEEE (2015). p. 1459–62.
30. Yu B, Zhan P, Lei M, Zhou F, Wang P. Food quality monitoring system based on smart contracts and evaluation models. *IEEE Access*. (2020) 8:12479–90. doi: 10.1109/ACCESS.2020.2966020
31. Prajwal A, Vaishali P, Sumit D. Food quality detection and monitoring system. In: *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Bhopal: IEEE (2020). p. 1–4.
32. Considine KM, Kelly AL, Fitzgerald GF, Hill C, Sleator RD. High-pressure processing-effects on microbial food safety and food quality. *FEMS Microbiol Lett*. (2008) 281:1–9. doi: 10.1111/j.1574-6968.2008.01084.x
33. Raveau R, Fontaine J, Lounès-Hadj Sahraoui A. Essential oils as potential alternative biocontrol products against plant pathogens and weeds: a review. *Foods*. (2020) 9:365. doi: 10.3390/foods9030365
34. Bastos DH, Saldanha LA, Catharino RR, Sawaya A, Cunha IB, Carvalho PO, et al. Phenolic antioxidants identified by ESI-MS from yerba mate (Ilex paraguariensis) and green tea (Camelia sinensis) extracts. *Molecules*. (2007) 12:423–32. doi: 10.3390/12030423
35. Anand G, Prakash L. IoT based novel smart refrigerator to curb food wastage. In: *2018 3rd International Conference on Contemporary Computing and Informatics (IC3I)*. Gurgaon: IEEE (2018). p. 268–72.
36. Chanthini B, Manivannan D, Umamakeswari A. Perishable food quality monitoring–an internet of things (IoT) approach. *Int J Pure Appl Math*. (2017) 115:63–66. Available online at: https://acadpubl.eu/jsi/2017-115-6-7/articles/7/10.pdf
37. Ilie AM, Vaccaro C. Design of a smart gas detection system in areas of natural gas storage. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX: IEEE (2017). p. 5954–7.
38. Jain P, Kushwaha R. Wireless gas sensor network for detection and monitoring of harmful gases in utility areas and industries. In: *2012 Sixth International Conference on Sensing Technology (ICST)*. Kolkata: IEEE (2012). p. 642–6.
39. Jovanović U, Jovanović I, Mančić D. Overview of temperature sensors for temperature measurement of PV modules. In: *2018 26th Telecommunications Forum (TELFOR)*. Belgrade: IEEE (2018). p. 1–8.
40. Benabdellah N, Bourhaleb M, Benazzi N, Dahbi S. Design of temperature and humidity sensors for an electronic nose used in rotten food. In: *2016 International Conference on Electrical and Information Technologies (ICEIT)*. Tangiers: IEEE (2016). p. 505–9.
41. Oltean M. *Fruit-360 Dataset Available*. (2021). Available online at: https://www.kaggle.com/moltean/fruits
42. *Fruit and Veggies Dataset Available*. (2021). Available online at: https://docs.google.com/spreadsheets/d/1Cz8N7z6tJ8FouYnbbn5shy2D71Ln5P0W/edit?usp=sharing&ouid=105439467837665352046&rtpof=true&sd=true

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sonwani, Bansal, Alroobaea, Baqasah and Hedabou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fpubh.2021.815631

A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Hypopharyngeal Squamous Cell Carcinomas: A Population-Based Study

JinKui Wang[1†], *XiaoZhu Liu*[2†], *Jie Tang*[3], *Qingquan Zhang*[4] *and Yuanyang Zhao*[5*]

[1] Chongqing Key Laboratory of Pediatrics, Department of Urology, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders (Chongqing), China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China, [2] Department of Cardiology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [3] Department of Epidemiology, Shenyang Medical College, Public Health School, Shenyang, China, [4] Department of Otorhinolaryngology and Head and Neck Surgery, Yuhuangding Hospital of Qingdao University, Yantai, China, [5] Department of Otolaryngology, Armed Police Hospital of Chongqing, Chongqing, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Abdul Rehman Javed, Air University, Pakistan
Gokul Yenduri, VIT University, India

***Correspondence:**
Yuanyang Zhao
zhaoy2014@outlook.com

[†]*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *15 November 2021*
Accepted: *20 December 2021*
Published: *13 January 2022*

Citation:
Wang J, Liu X, Tang J, Zhang Q and Zhao Y (2022) A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Hypopharyngeal Squamous Cell Carcinomas: A Population-Based Study. Front. Public Health 9:815631. doi: 10.3389/fpubh.2021.815631

Background: Hypopharyngeal squamous cell carcinomas (HPSCC) is one of the causes of death in elderly patients, an accurate prediction of survival can effectively improve the prognosis of patients. However, there is no accurate assessment of the survival prognosis of elderly patients with HPSCC. The purpose of this study is to establish a nomogram to predict the cancer-specific survival (CSS) of elderly patients with HPSCC.

Methods: The clinicopathological data of all patients from 2004 to 2018 were downloaded from the SEER database. These patients were randomly divided into a training set (70%) and a validation set (30%). The univariate and multivariate Cox regression analysis confirmed independent risk factors for the prognosis of elderly patients with HPSCC. A new nomogram was constructed to predict 1-, 3-, and 5-year CSS in elderly patients with HPSCC. Then used the consistency index (C-index), the calibration curve, and the area under the receiver operating curve (AUC) to evaluate the accuracy and discrimination of the prediction model. Decision curve analysis (DCA) was used to assess the clinical value of the model.

Results: A total of 3,172 patients were included in the study, and they were randomly divided into a training set (N = 2,219) and a validation set (N = 953). Univariate and multivariate analysis suggested that age, T stage, N stage, M stage, tumor size, surgery, radiotherapy, chemotherapy, and marriage were independent risk factors for patient prognosis. These nine variables are included in the nomogram to predict the CSS of patients. The C-index for the training set and validation was 0.713 (95% CI, 0.697–0.729) and 0.703 (95% CI, 0.678–0.729), respectively. The AUC results of the training and validation set indicate that this nomogram has good accuracy. The calibration curve indicates that the observed and predicted values are highly consistent. DCA indicated that the nomogram has a better clinical application value than the traditional TNM staging system.

Conclusion: This study identified risk factors for survival in elderly patients with HPSCC. We found that age, T stage, N stage, M stage, tumor size, surgery, radiotherapy, chemotherapy, and marriage are independent prognostic factors. A new nomogram for predicting the CSS of elderly HPSCC patients was established. This model has good clinical application value and can help patients and doctors make clinical decisions.

Keywords: nomogram, HPSCC, elderly patients, cancer-specific survival, online application

BACKGROUND

Hypopharyngeal carcinoma, which can occur in the pyriform sinuses bilaterally, the posterior and lateral pharyngeal walls, and the postcricoid region (1), accounts for approximately 3–5% of all head and neck carcinoma (2, 3). According to the Surveillance, Epidemiology, and End Results database (SEER), from 1973 to 2008, elderly over 65 years old with head and neck cancer accounted for about 47% of the total head and neck cancer in the United States (4). Most hypopharyngeal cancers are squamous cell carcinomas (HPSCC), which are less common than other head and neck malignancies and account for about 3–5% of squamous cell carcinomas of the head and neck. The incidence rate of cancer in the elderly increases with age and is higher in people over 65 (5, 6). So as aging intensifies, the number of elderly with HPSCC is expected to rise in the future. The prognosis of HPSCC is poor, with a 5-year survival rate being 25–47% (7, 8). Elderly patients are a heterogeneous population with their performance status (9). Comorbidity, organ dysfunction and immunosenescence as adverse factors affect the treatment options of elderly patients and let patients experience more treatment-related toxicity and poor prognosis (4, 9).

Gang et al. found that T3-4 stage, N2-3 stage, and poor tumor differentiation were risk factors for the Development of vascular invasion in HPSCC and reduced patients' overall survival (OS) and cancer-specific survival (CSS) (10). Iuchi et al. explored the modified Glasgow prognostic score as a beneficial prognostic factor for patients with HPSCC (11). Zhongyang et al. constructed a nomogram to predict survival in HPSCC and found that age, race, TNM stage, surgery, radiotherapy, chemotherapy, and marriage were independent risk factors for patients (12). Although there are nomograms to predict the prognosis of patients with head and neck cancer and even hypopharyngeal squamous cell carcinoma, there is no study to specifically construct the nomogram for elderly patients with HPSCC to predict the CSS. As a high-risk cancer population, the elderly have particular survival and prognostic factors. Therefore, a special nomogram can improve the accuracy and practical value of the prediction model.

Artificial intelligence has been widely used in human public health. Javed et al. applied Cognitive Assessment of Smart Home Resident to detect individuals with early cognitive impairment (13). Abbas et al. used a new machine-learning algorithm, BCD-WERT, to detect breast cancer, which has better accuracy than the traditional machine learning method (14).

At present, many prediction models are applied to human health, such as software metrics by Grey Wolf Optimization proposed by Yenduri et al. (15). The traditional line chart is also a commonly used prediction model. Nomogram is a visually reliable statistical tool to calculate conveniently and present the survival probability of a certain time-point for a particular patient based on risk factors included (16–18). However, there have been few Analyses of influencing factors and survival trends for hypopharyngeal cancer within the elderly population. Accurate prediction of the survival of elderly patients with HPSCC can help doctors answer patient inquiries. It also can help doctors and patients to develop treatment and follow-up strategies. We aim to explore the risk factors for CSS in elderly patients with HPSCC and construct a nomogram for predicting CSS to help doctors and patients make treatment and follow-up decisions.

PATIENTS AND METHODS

Data Source and Data Extraction

The clinicopathological data of all patients with HPSCC from 2004 to 2018 were downloaded from the SEER database. Including age, gender, race, year of diagnosis, marriage, tumor location, tumor size, histopathological grade, histological type, TNM stage, surgery, radiotherapy, chemotherapy, and other information. Since the information in the SEE database is public and anonymous, our study does not require ethical approval and patient informed consent. Our research methods follow the guidelines of the SEER database.

Inclusion Criteria

(1) age $\geq$ 60 years; (2)primary tumor site is hypopharynx; (3) histopathological type is squamous cell carcinoma.

Exclusion Criteria

(1) Survival time is $<$1 month; (2) tumor size is unknown; (3) TNM staging is unknown; (4) surgical method is unknown. The age of the patients was divided into three groups, including 60–69 years old, 70–79 years old, $\geq$80 years old. The race included white, black, and other races (American Indian/AK Indian, Asian/Pacific Islander). The tumor site was divided into the pyriform sinus and others, including the postcricoid region, aryepiglottic fold, hypopharyngeal, posterior wall of the hypopharynx. The histological grades of tumors were grade I-IV, well-differentiated, moderately differentiated, poorly differentiated, and undifferentiated. The tumor size is the most significant part of the tumor diameter, divided into $\leq$2 cm, 2–4 cm, $\geq$4 cm. The screening process for all patients is shown in **Figure 1**.

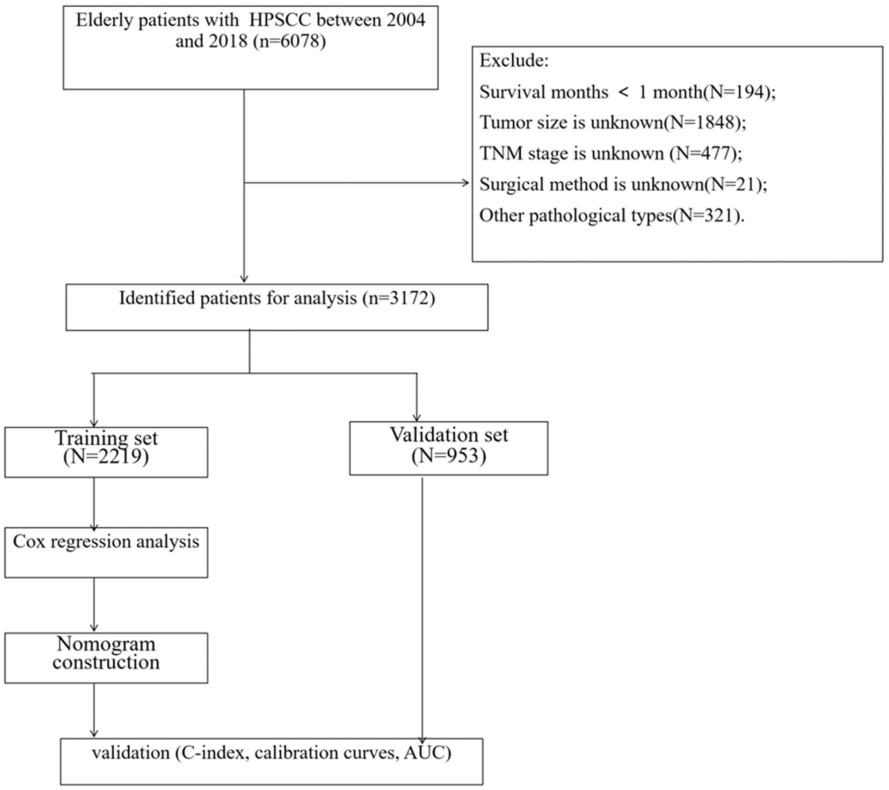


FIGURE 1 | The flowchart of including and dividing patients.

Nomogram Construction and Validation

All patients were divided into a training set (70%) and a validation set (30%). The prognostic factors of patients need to be clarified. Cox proportional hazard regression model is the most commonly used regression model to explore the prognosis of cancer patients. The patients in the training set were subjected to univariate Cox regression analysis to screen out the prognostic factors of HPSCC. The confirmed related elements were included in the multivariate Cox regression analysis to screen out independent risk factors for HPSCC. A new nomogram was constructed to predict 1-, 3-, and 5-year CSS in elderly patients with HPSCC. Then, the consistency index (C-index) and the area under the receiver operating curve (AUC) were used to evaluate the accuracy and discrimination of the prediction model. The calibration curve of 1000 bootstrap weight samples was used to verify the accuracy of the nomogram.

Clinical Utility

Decision curve analysis (DCA) was used to evaluate the clinical value of the nomogram. It is a new algorithm to consider the application value of the model based on various risk thresholds (19). At the same time, the clinical significance of nomogram and traditional TNM staging system was compared based on DCA. Subsequently, the patient's survival risk value was calculated based on the nomogram, and then the patients were divided into high-risk and low-risk groups based on the cut-off value. Kaplan-Meier (K-M) curve and log-rank test were used to compare the difference in survival of patients in each group.

Statistical Analysis

All statistics were performed using SPSS 26.0 and R software 4.1.0. The count data were described by frequency (%) and analyzed by chi-square and non-parametric U-test. Univariate and multivariate Cox regression analyses were used to analyze the prognostic factors of patient survival. Kaplan-Meier curve and log-rank test were used to compare the survival differences of patients. A $P < 0.05$ was considered statistically significant.

RESULTS

Clinical features

A total of 3,172 elderly patients with HPSCC were included. These patients were randomly divided into a training set ($N = 2,219$) and a validation set ($N = 953$). The clinicopathological information of the patients is shown in **Table 1**. Among them, 1,578 cases (49.7%) were 60–69 years old, 1,122 cases (35.4%) were 70–79 years old, and 472 cases (14.9%) were ≥80 years old. There were 2,513 whites (79.2%), 2,558 men (80.6%), and

TABLE 1 | Clinicopathological characteristics of children with HPSCC.

	Total	Training set	Validation set	
	N = 3,172	*N* = 2,219	*N* = 953	*p*
Age				0.374
60–69	1,578 (49.7%)	1,087 (49.0%)	491 (51.5%)	
70–79	1,122 (35.4%)	801 (36.1%)	321 (33.7%)	
≥80	472 (14.9%)	331 (14.9%)	141 (14.8%)	
Race				0.536
white	2,513 (79.2%)	1,764 (79.5%)	749 (78.6%)	
black	412 (13.0%)	279 (12.6%)	133 (14.0%)	
other	247 (7.79%)	176 (7.93%)	71 (7.45%)	
Sex				0.767
Male	2,558 (80.6%)	1,793 (80.8%)	765 (80.3%)	
Female	614 (19.4%)	426 (19.2%)	188 (19.7%)	
Year of diagnosis				0.052
2004-2008	929 (29.3%)	660 (29.7%)	269 (28.2%)	
2009-2013	1,129 (35.6%)	760 (34.2%)	369 (38.7%)	
2014-2018	1,114 (35.1%)	799 (36.0%)	315 (33.1%)	
Grade				0.744
I	124 (3.91%)	88 (3.97%)	36 (3.78%)	
II	1,259 (39.7%)	869 (39.2%)	390 (40.9%)	
III	1,063 (33.5%)	749 (33.8%)	314 (32.9%)	
IV	29 (0.91%)	23 (1.04%)	6 (0.63%)	
Unknown	697 (22.0%)	490 (22.1%)	207 (21.7%)	
T				0.184
T1	408 (12.9%)	288 (13.0%)	120 (12.6%)	
T2	1,138 (35.9%)	819 (36.9%)	319 (33.5%)	
T3	619 (19.5%)	416 (18.7%)	203 (21.3%)	
T4	1,007 (31.7%)	696 (31.4%)	311 (32.6%)	
N				0.328
N0	1,117 (35.2%)	783 (35.3%)	334 (35.0%)	
N1	661 (20.8%)	451 (20.3%)	210 (22.0%)	
N2	1,278 (40.3%)	910 (41.0%)	368 (38.6%)	
N3	116 (3.66%)	75 (3.38%)	41 (4.30%)	
M				0.405
M0	2,949 (93.0%)	2,069 (93.2%)	880 (92.3%)	
M1	223 (7.03%)	150 (6.76%)	73 (7.66%)	
Marital				0.828
No	1,582 (49.9%)	1,110 (50.0%)	472 (49.5%)	
Married	1,590 (50.1%)	1,109 (50.0%)	481 (50.5%)	
Tumor size				0.041
≤2	747 (23.5%)	532 (24.0%)	215 (22.6%)	
2–4	1,547 (48.8%)	1,102 (49.7%)	445 (46.7%)	
≥4	878 (27.7%)	585 (26.4%)	293 (30.7%)	
Primary site				0.627
Pyriform sinus	1,635 (51.5%)	1,137 (51.2%)	498 (52.3%)	
Other	1,537 (48.5%)	1,082 (48.8%)	455 (47.7%)	
Chemotherapy				0.889
No/Unknown	1,309 (41.3%)	918 (41.4%)	391 (41.0%)	
Yes	1,863 (58.7%)	1,301 (58.6%)	562 (59.0%)	
Radiation				0.442
No/Unknown	779 (24.6%)	554 (25.0%)	225 (23.6%)	
Yes	2,393 (75.4%)	1,665 (75.0%)	728 (76.4%)	
Surgery				0.208
No	2,336 (73.6%)	1,649 (74.3%)	687 (72.1%)	
Yes	836 (26.4%)	570 (25.7%)	266 (27.9%)	

TABLE 2 | Univariate and multivariate analyses of CSS in training set.

	Univariate			Multivariate		
	HR	95%CI	P	HR	95%CI	P
Age						
60–69						
70–79	1.032	0.905–1.177	0.634	1.109	0.97–1.269	0.13
≥80	1.716	1.456–2.023	<0.001	1.679	1.414–1.994	<0.001
Sex						
Male						
Female	0.93	0.8–1.08	0.361			
Year of diagnosis						
2004–2008						
2009–2013	1.001	0.871–1.149	0.991			
2014–2018	0.895	0.767–1.044	0.159			
Race						
white						
black	1.404	1.182–1.668	<0.001			
other	1.113	0.893–1.387	0.342			
Primary site						
Pyriform sinus						
Other	1.1	0.98–1.24	0.105			
Grade						
I						
II	0.913	0.68–1.227	0.546			
III	0.91	0.676–1.225	0.534			
IV	0.681	0.354–1.31	0.249			
Unknown	0.915	0.672–1.246	0.573			
T						
T1						
T2	1.328	1.065–1.655	0.012	1.264	0.965–1.657	0.089
T3	2.118	1.679–2.674	<0.001	1.635	1.204–2.221	0.002
T4	2.669	2.153–3.309	<0.001	2.196	1.644–2.935	<0.001
N						
N0						
N1	1.059	0.891–1.259	0.514	1.166	0.975–1.394	0.093
N2	1.492	1.302–1.709	<0.001	1.574	1.36–1.821	<0.001
N3	2.647	1.955–3.584	<0.001	2.296	1.68–3.137	<0.001
M						
M0						
M1	3.11	2.56–3.78	<0.001	1.992	1.622–2.447	<0.001
Surgery						
No						
Yes	0.67	0.58–0.77	<0.001	0.508	0.434–0.594	<0.001
Radiation						
No/Unknown						
Yes	0.54	0.47–0.61	<0.001	0.513	0.441–0.597	<0.001
Chemotherapy						
No/Unknown						
Yes	0.8	0.71–0.91	<0.001	0.705	0.609–0.815	<0.001
Tumor size						
≤2						
2–4	1.357	1.16–1.587	<0.001	1.028	0.84–1.259	0.786
≥4	2.236	1.891–2.644	<0.001	1.436	1.136–1.815	0.002
Marital						
No						
Married	0.66	0.59–0.75	<0.001	0.726	0.644–0.819	<0.001

1,590 married (50.1%). The histological tumor grades were 124 (3.91%), 1,259 (39.7%), 1,063 (33.5%), and 29 (0.91%) patients with tumor histological grades I, II, III, and IV, respectively. There were 1,635 cases (51.5%) whose tumor site was pyriform sinus. There were 747 cases (23.5%), 1,547 cases (48.8%), and 878 cases (27.7%) with tumor sizes of $\leq$2 cm, 2–4 cm, and $\geq$4 cm, respectively. There were 2,336 cases (73.6%) that underwent surgery, 2,393 cases (75.4%) underwent radiotherapy, and 1,863 cases (58.7%) underwent chemotherapy. There is no significant difference between the clinical-pathological information of the training set and the validation set.

Univariate and Multivariate Cox Regression Analysis

Univariate Cox regression analysis was used to screen the prognostic factors of HPSCC in elderly patients, including age, race, T stage, N stage, M stage, surgery, radiotherapy, chemotherapy, tumor size, and marriage. Multivariate analysis found that age, T stage, N stage, M stage, surgery, radiotherapy, chemotherapy, tumor size, and marriage were independent risk factors for the prognosis of patients. The results of univariate and multivariate factors are shown in **Table 2**. The results showed that older patients had a worse prognosis. The higher the TNM stage, the worse prognosis. In addition, patients with larger tumors have a worse prognosis. Patients have higher survival rates after surgery, radiotherapy, and chemotherapy. Marital status is a favorable factor for the survival of patients.

Nomogram Construction and Validation

The multivariate Cox regression analysis identified the independent risk factors used to construct a nomogram to predict elderly swallowing at 1-, 3-, and 5-year CSS (**Figure 2**). The nomogram indicated that TNM staging is still the most significant factor affecting the prognosis of patients. In addition, surgery and radiotherapy are the most important prognostic factors among other variables. They are then followed by age, tumor size, chemotherapy, and marriage. The C-index of the training set and the validation set were 0.713 (95% CI, 0.697–0.729) and 0.703 (95% CI, 0.678–0.729), respectively, indicating that the nomogram has good accuracy. The calibration curves of the training and validation sets are consistent with the 45° diagonal, meaning that the observed value is consistent with the predicted value (**Figure 3**). The 1-, 3-, 5-year AUC of the training set and validation set indicates that the nomogram has better accuracy than TNM staging (**Figure 4**).

Clinical Application of the Nomogram

The DCA of the training set and the validation set still suggested that the nomogram has a better clinical application value than the traditional TNM staging (**Figure 5**). According to the constructed nomogram, each patient received a risk score. According to the score, patients were divided into high-risk groups (total score $\geq$198.9) and low-risk groups (total score $\leq$198.9). The K-M curves of the validation set and the training set proved that the survival of patients in each risk group is significantly different (**Figure 6**). The K-M curve result proved the discriminative power of the nomogram prediction model. We analyzed the difference in survival of patients in other risk groups under different treatments. In the low-risk group, we found no significant difference in survival between patients with and without surgery (**Figure 7A**). Similarly, there was no difference in survival between patients who have been treated with radiation and those who have not (**Figure 7C**). However, patients in the low-risk group who received chemotherapy had a lower survival rate than those who did not (**Figure 7B**). In the high-risk group, patients with surgery have a higher survival rate than patients without surgery (**Figure 7D**). Patients with radiation therapy had a higher survival rate than patients without radiation therapy (**Figure 7F**). Patients with chemotherapy had a higher survival rate than patients without chemotherapy (**Figure 7E**).

Online Application for CSS Prediction

Based on this nomogram, we developed a network calculator at https://prediction-app.shinyapps.io/DynNomapp/. Enter the clinicopathological characteristics of the patient on this website to get the patient's predictive CSS. This website is straightforward and practical, convenient for patients and patients.

DISCUSSION

HPSCC is one of the most malignant head and neck carcinomas with a poor prognosis (1, 20). More than 50% of patients with HPSCC were usually diagnosed at a more advanced stage (21). Various factors lead to this result, such as minor symptoms in the early stages, early nodal metastasis, and poor differentiation (22, 23). The prognosis of HPSCC is one of the worst for head and neck tumors. The reason for this is that the hypopharynx has a rich lymphatic network, and cancer cells can metastasize through the lymphatic system at a very early stage of the disease (24, 25). Also, early diagnosis of HPSCC is often tricky as there are almost no specific clinical symptoms in the early stages. As a result, most patients are already in the advanced stages of cancer by the time the disease is diagnosed (26, 27). Elderly patients with HPSCC are a high-risk group and should receive special attention. An accurate predictive assessment can provide advice to patients and provide a basis for individualized treatment.

TNM staging is the primary tumor staging system, which is an essential factor in evaluating cancer prognosis (28). However, precise evaluation of TNM staging can only be obtained after surgery (29), and prognostic heterogeneity generally exists among patients with the same TNM stage (30). Hypopharynx is rich in lymphoid tissue and prone to mucosal infiltrative metastases (31). Sometimes, physical examination of lymph nodes is inaccurate, and small metastases cannot be detected by imaging (32), limiting prognostic evaluation accuracy by TNM staging. It should be noted that the prognosis of cancer is affected by many factors, especially in the

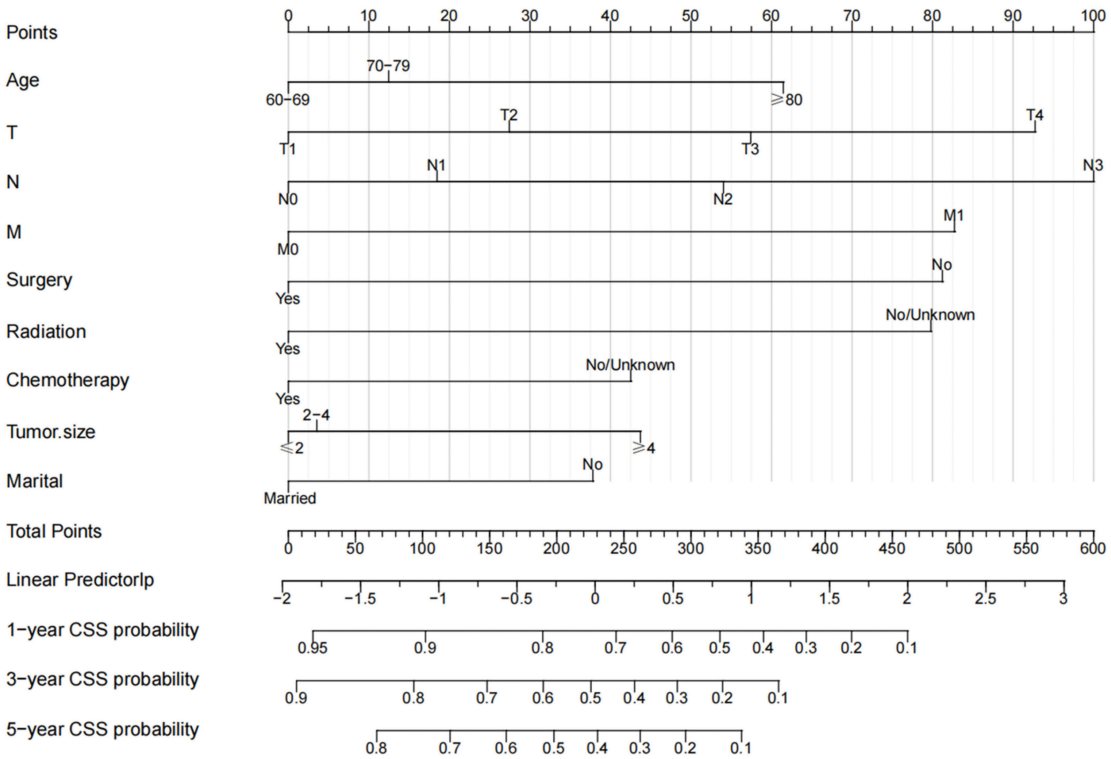


FIGURE 2 | Nomogram for 1-, 3-, and 5-year CSS of elderly patients with HPSCC.

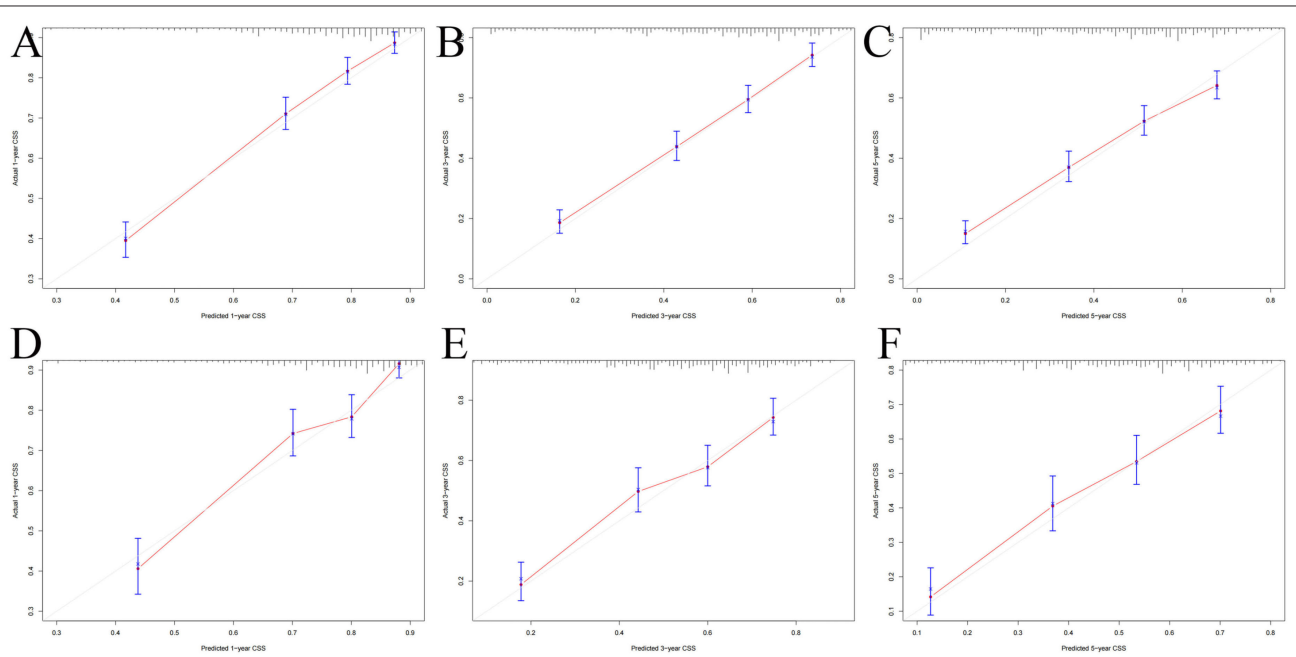


FIGURE 3 | Calibration curves of the nomogram. **(A–C)** For 1-, 3-, and 5-year CSS in the training set; **(D–F)** For 1-, 3-, and 5-year CSS in the validation set.

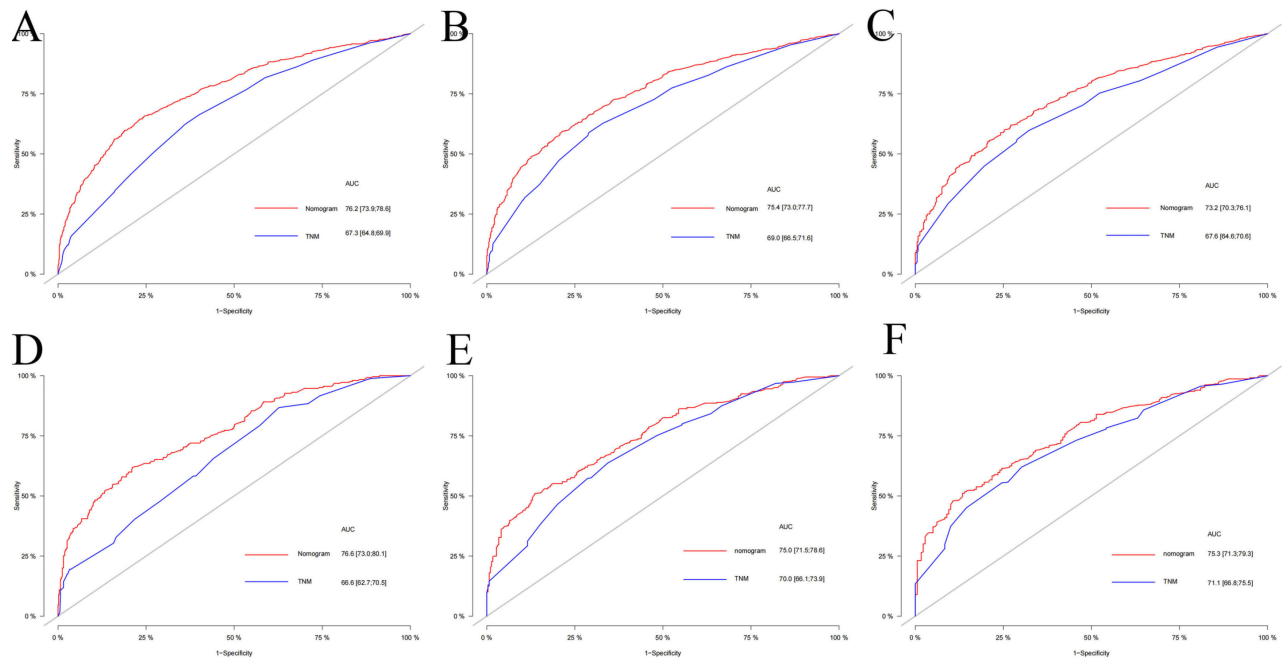


FIGURE 4 | The AUC of nomogram and TNM stage of 1-, 3- and 5-year in the training set **(A–C)** and validation set **(D–F)**.

elderly. Previous studies showed that demographics, clinicopathological parameters, and treatment factors determine survival outcomes for hypopharyngeal carcinoma (33). In our research, the multivariable Cox regression analysis revealed that age at diagnosis, marital status, tumor size, TNM stage, surgery, radiotherapy, and chemotherapy is associated with the prognosis of HPSCC.

This study found that age was still a prognostic factor for patients and that survival was lower in patients of higher age. Age as a prognostic factor is consistent with previous findings (34, 35). It is worth noting that married patients have a higher survival rate than those who are not married. Not surprisingly, marriage gives patients more emotional comfort and better financial support (36). In addition, various treatments have been shown to impact the survival of elderly patients with HPSCC significantly. Currently, radical laryngectomy is gradually becoming laryngeal preservation surgery as the procedure improves (37). Our study found that patients who had surgery still had a higher survival rate than those who did not have surgery. In addition, radiotherapy and chemotherapy offer patients the option of laryngeal preservation surgery (38). Our study showed that patients had significantly better survival rates with radiotherapy and chemotherapy.

In addition, patients were divided into high-risk and low-risk groups based on the scores of the line graphs. We found that surgery and radiotherapy for patients in the low-risk group did not significantly improve patient survival. However, chemotherapy appeared to improve the survival prognosis of elderly patients significantly. Therefore, chemotherapy should be routinely recommended for elderly patients with HPSCC in the low-risk group. For patients in the high-risk group, surgery, radiotherapy, and chemotherapy are all necessary as they effectively improve patient survival.

Although nomograms have been constructed to predict OS and CSS in hypopharyngeal cancer, no nomograms have been explicitly built to expect CSS in elderly patients with HPSCC (39, 40). Due to the population-specific nature of the elderly, there may be prognostic factors that differ from those of other populations. Therefore, based on these factors, we developed a nomogram to predict the prognosis of the elderly with HPSCC. The calibration curve and C-index of the prediction model suggest good accuracy and reliability. Based on the AUC and DCA results, this new nomogram is superior to conventional TNM staging.

Our research also has some defects. First, there is no information about the comorbidity of the elderly in the SEER database, such as coronary heart disease, hypertension, and diabetes. These complications affect the choice of treatment options. We cannot incorporate these factors into the prognosis analysis, so the accuracy is limited. However, we included essential patient factors such as age, TNM stage, surgery, radiotherapy, chemotherapy, etc. These factors are critical determinants of patient survival. Second, as this study was a retrospective case

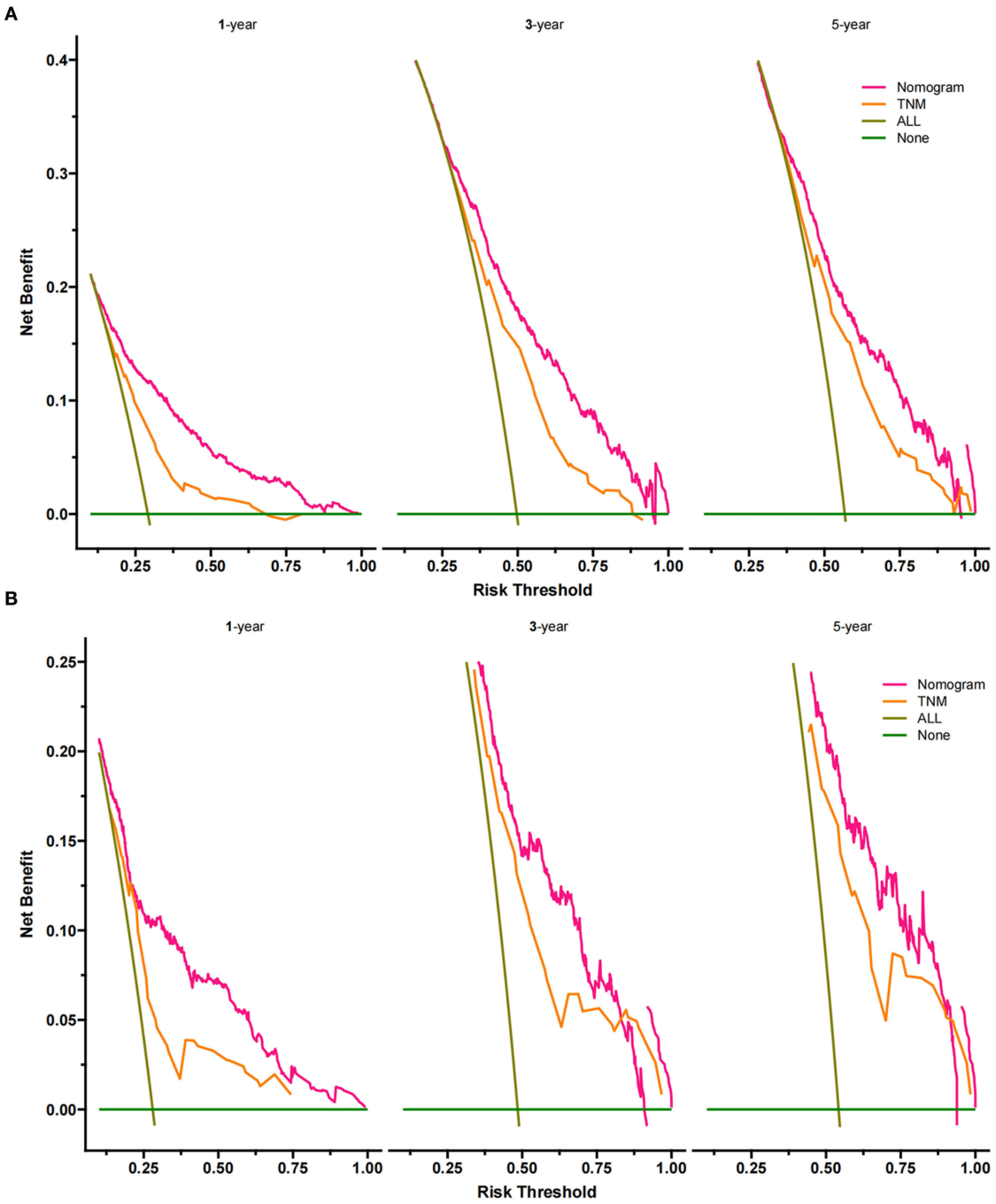


FIGURE 5 | Decision curves of the nomogram predicting CSS in the training set **(A)** and validation set **(B)**. The Y-axis represents net income, and the X-axis represents threshold probability. The green line means no patients died, and the dark green line means all patients died. When the threshold probability is between 25 and 90%, the net benefit of the model exceeds all deaths or none.

study, a selection bias was difficult to adjust for and may have contributed to some errors in the results. Finally, although we have conducted split internal validation on the SEER database, we lack multi-center external validation, especially data validation from China. This limits the application of the nomogram in areas other than the United States. We intend to further validate the accuracy of the prediction model through external patient data. The data were collected prospectively according to the variables of this model, and the patients were further

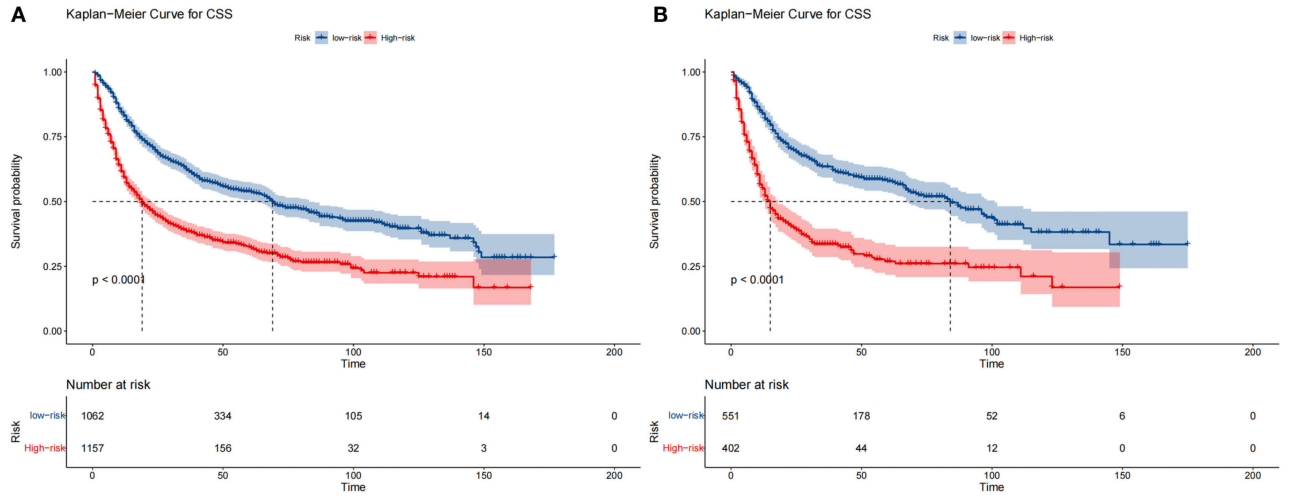


FIGURE 6 | Kaplan–Meier curves of CSS for patients in the low- and high-risk groups in the training set **(A)** and validation set **(B)**.

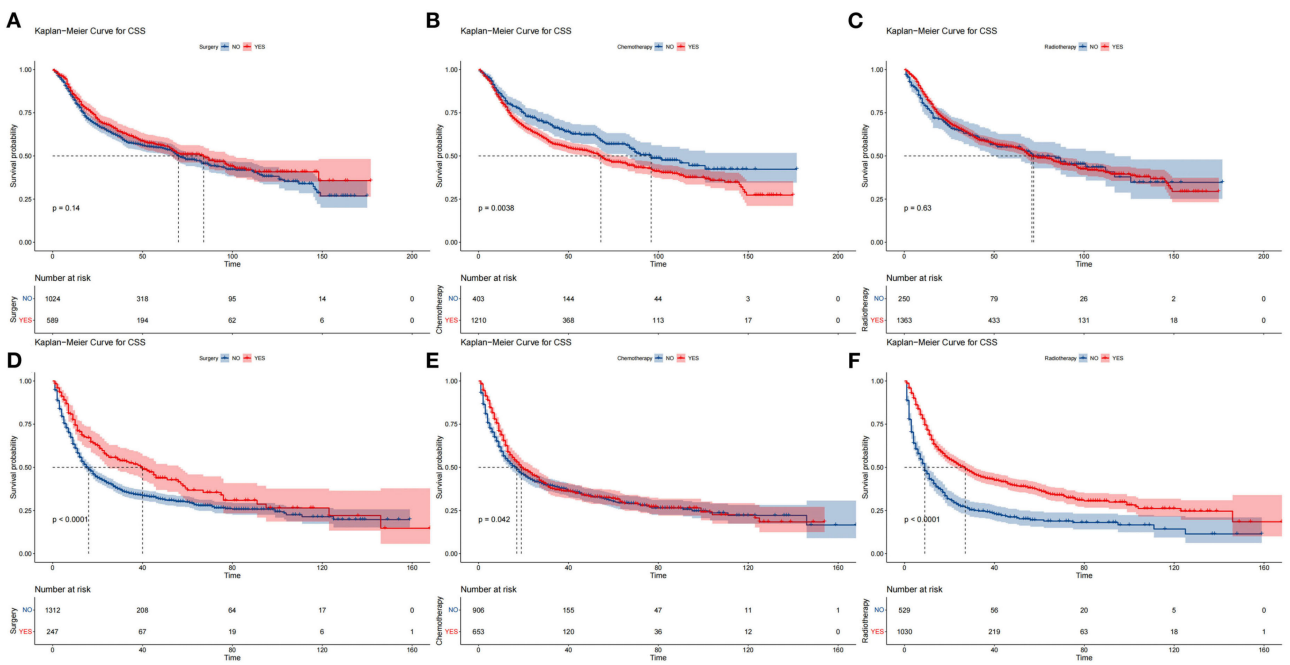


FIGURE 7 | Comparison of different surgical methods of Kaplan–Meier curves. Patients with or without surgery, chemotherapy, and radiation in the low-risk group **(A–C)** and high-risk group **(D–F)**.

followed up. Then the patients were included in the model for external validation.

CONCLUSION

In conclusion, we established and validated a new nomogram to predict the prognosis of elderly patients with HPSCC. Through this model, clinicians can estimate the survival rate of elderly patients with HPSCC more accurately, which plays an essential role in guiding the choice of clinical treatment.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov/.

AUTHOR CONTRIBUTIONS

JW, YZ, and XL contributed to the conception and design. JW, JT, and XL collected and analyzed the data. JW, QZ, and JT drew the figures and tables. JT, YZ, and JW wrote the draft. XL and JW contributed to manuscript writing and revision. All authors approved the final manuscript.

REFERENCES

1. Hinerman RW, Amdur RJ, Mendenhall WM, Villaret DB, Robbins KT. Hypopharyngeal carcinoma. *Curr Treat Options Oncol.* (2002) 3:41–9. doi: 10.1007/s11864-002-0040-1
2. Tsai YT, Chen WC, Chien CY, Hsu CM, Lee YC, Tsai MS, Lin MH, Lai CH, Chang KP. Treatment patterns and survival outcomes of advanced hypopharyngeal squamous cell carcinoma. World J Surg Oncol. (2020) 18:82. doi: 10.1186/s12957-020-01866-z
3. Kajanti M, Mäntylä M. Carcinoma of the hypopharynx. A retrospective analysis of the treatment results over a 25-year period. *Acta Oncol.* (1990) 29:903–7. doi: 10.3109/02841869009096387
4. VanderWalde NA, Fleming M, Weiss J, Chera BS. Treatment of older patients with head and neck cancer: a review. *Oncologist.* (2013) 18:568–78. doi: 10.1634/theoncologist.2012-0427
5. Muramatsu N, Akiyama H. Japan: super-aging society preparing for the future. *Gerontologist.* (2011) 51:425–32. doi: 10.1093/geront/gnr067
6. Yasui M, Sakaguchi M, Jikuya R, Tsutsumi S, Tatenuma T, Noguchi G, et al. Comparative effectiveness of surgery and radiotherapy for survival of patients with clinically localized prostate cancer: A population-based coarsened exact matching retrospective cohort study. *Oncol Lett.* (2020) 20:150. doi: 10.3892/ol.2020.12013 Epub 2020 Aug 24. PMID: 32934718; PMCID: PMC7475640.
7. Bu M, Liu X, Liu X, Xu W. Upregulation of fascin-1 is involved in HIF-1α-dependent invasion and migration of hypopharyngeal squamous cell carcinoma. *Int J Oncol.* (2019) 55:488–98. doi: 10.3892/ijo.2019.4827
8. Bradley PJ. Epidemiology of hypopharyngeal cancer. *Adv Otorhinolaryngol.* (2019) 83:1–14. doi: 10.1159/000492299
9. Lee L, Cheung WY, Atkinson E, Krzyzanowska MK. Impact of comorbidity on chemotherapy use and outcomes in solid tumors: a systematic review. *J Clin Oncol.* (2011) 29:106–17. doi: 10.1200/JCO.2010.31.3049
10. Gang G, Xinwei C, LiXiao C, Yu Z, Cheng Z, Pin D. Risk factors of lymphovascular invasion in hypopharyngeal squamous cell carcinoma and its influence on prognosis [published online ahead of print, 2021 Jun 2]. *Eur Arch Otorhinolaryngol.* (2021) 10.1007/s00405-021-06906-2. doi: 10.1007/s00405-021-06906-2
11. Iuchi H, Kyutoku T, Ito K, Matsumoto H, Ohori J, Yamashita M. Impacts of inflammation-based prognostic scores on survival in patients with hypopharyngeal squamous cell carcinoma. *OTO Open.* (2020) 4:2473974X20978137. doi: 10.1177/2473974X20978137
12. Lin Z, Lin H, Lin C. Dynamic prediction of cancer-specific survival for primary hypopharyngeal squamous cell carcinoma. *Int J Clin Oncol.* (2020) 25:1260–9. doi: 10.1007/s10147-020-01671-4
13. Javed AR, Fahad LG, Farhan AA, Abbas S, Srivastava G, Parizi RM, et al. Automated cognitive health assessment in smart homes using machine learning. *Sustainable Cities and Society.* (2021) 65:102572. doi: 10.1016/j.scs.2020.102572
14. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci.* (2021) 7:e390. doi: 10.7717/peerj-cs.390
15. Yenduri G, Naralasetti V. A nonlinear weight-optimized maintainability index of software metrics by Grey Wolf Optimization. *Int J Swarm Intell Res.* (2021) 12:1–21. doi: 10.4018/IJSIR.2021040101
16. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7
17. Karakiewicz PI, Briganti A, Chun FK, Trinh QD, Perrotte P, Ficarra V, et al. Multi-institutional validation of a new renal cancer-specific survival nomogram. *J Clin Oncol.* (2007) 25:1316–22. doi: 10.1200/JCO.2006.06.1218 PMID: 17416852.
18. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol.* (2015) 33:861–9. doi: 10.1200/JCO.2014.56.6661 Epub 2015 Jan 26. PMID: 25624438.
19. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* (2008) 8:53. doi: 10.1186/1472-6947-8-53
20. Newman JR, Connolly TM, Illing EA, Kilgore ML, Locher JL, Carroll WR. Survival trends in hypopharyngeal cancer: a population-based review. *Laryngoscope.* (2015) 125:624–9. doi: 10.1002/lary.24915
21. Liu J, Zhu W, Li Z, Cai G, Wang J, Tang Q, et al. Proteomic analysis of hypopharyngeal and laryngeal squamous cell carcinoma sheds light on differences in survival. *Sci Rep.* (2020) 10:19459. doi: 10.1038/s41598-020-76626-w
22. Zhou J, Li Y, Wei D, Qian Y, Li W, Liu D, et al. Overall survival with and without laryngeal function preservation in 580 patients with hypopharyngeal squamous cell carcinoma. *Oncol Rep.* (2015) 34:3196–202. doi: 10.3892/or.2015.4313
23. Wang Z, Wei P, Wei D, Cao S, Liu H, Chen L, et al. Effect of up-regulation of circMATR3 on the proliferation, metastasis, progression and survival of hypopharyngeal carcinoma. *J Cell Mol Med.* (2020) 24:4687–97. doi: 10.1111/jcmm.15134
24. Belcher R, Hayes K, Fedewa S, Chen AY. Current treatment of head and neck squamous cell cancer. *J Surg Oncol.* (2014) 110:551–74. doi: 10.1002/jso.23724
25. Chung EJ, Lee SH, Baek SH, Park IS, Cho SJ, Rho YS. Pattern of cervical lymph node metastasis in medial wall pyriform sinus carcinoma. *Laryngoscope.* (2014) 124:882–7. doi: 10.1002/lary.24299
26. Kim SY, Rho YS, Choi EC, Kim MS, Woo JH, Lee DH, et al. Clinicopathological factors influencing the outcomes of surgical treatment in patients with T4a hypopharyngeal cancer. *BMC Cancer.* (2017) 17:904. doi: 10.1186/s12885-017-3880-6
27. Kotwall C, Sako K, Razack MS, Rao U, Bakamjian V, Shedd DP. Metastatic patterns in squamous cell cancer of the head and neck. *Am J Surg.* (1987) 154:439–42. doi: 10.1016/0002-9610(89)90020-2
28. Wu H, Mu X, Liu L, Wu H, Hu X, Chen L, et al. Bone marrow mesenchymal stem cells-derived exosomal microRNA-193a reduces cisplatin resistance of non-small cell lung cancer cells via targeting LRRC1. *Cell Death Dis.* (2020) 11:801. doi: 10.1038/s41419-020-02962-4
29. Zhang J, Ding Y, Wang W, Lu Y, Wang H, Wang H, et al. Combining the Fibrinogen/Albumin Ratio and Systemic Inflammation Response Index Predicts Survival in Resectable Gastric Cancer. *Gastroenterol Res Pract.* (2020) 2020:3207345. doi: 10.1155/2020/3207345
30. Zhai Z, Zhu ZY, Zhang Y, Yin X, Han BL, Gao JL, et al. Prognostic significance of Borrmann type combined with vessel invasion status in advanced gastric cancer. *World J Gastrointest Oncol.* (2020) 12:992–1004. doi: 10.4251/wjgo.v12.i9.992
31. Wu Z, Deng XY, Zeng RF, Su Y, Gu MF, Zhang Y, et al. Analysis of risk factors for retropharyngeal lymph node metastasis in carcinoma of the hypopharynx. *Head Neck.* (2013) 35:1274–7. doi: 10.1002/hed.23112
32. Bluemel C, Herrmann K, Giammarile F, Nieweg OE, Dubreuil J, Testori A, et al. practice guidelines for lymphoscintigraphy and sentinel lymph node biopsy in melanoma. *Eur J Nucl Med Mol Imaging.* (2015) 42:1750–66. doi: 10.1007/s00259-015-3135-1

33. Tian S, Li Q, Li R, Chen X, Tao Z, Gong H, et al. Development and validation of a prognostic nomogram for hypopharyngeal carcinoma. *Front Oncol.* (2021) 11:696952. doi: 10.3389/fonc.2021.696952
34. Kim YJ, Lee R. Surgery vs. radiotherapy for locally advanced hypopharyngeal cancer in the contemporary era: A population-based study. *Cancer Med.* (2018) 7:5889–900. doi: 10.1002/cam4.1811
35. Psychogios G, Zenk J. Larynxerhalt durch Induktionschemotherapie bei Hypopharynxkarzinomen möglich [Laryngeal preservation with induction chemotherapy for hypopharyngeal squamous cell carcinoma]. *Strahlenther Onkol.* (2013) 189:1058–9. doi: 10.1007/s00066-013-0476-1
36. Zhou H, Zhang Y, Song Y, Tan W, Qiu Z, Li S, et al. Marital status is an independent prognostic factor for pancreatic neuroendocrine tumors patients: An analysis of the Surveillance, Epidemiology, and End Results (SEER) database. *Clin Res Hepatol Gastroenterol.* (2017) 41:476–86. doi: 10.1016/j.clinre.2017.02.008
37. Day D, Hansen AR, Siu LL. Hypopharyngeal cancer: looking back, moving forward. *Curr Oncol.* (2016) 23:221–2. doi: 10.3747/co.23.3242
38. Lefebvre JL, Andry G, Chevalier D, Luboinski B, Collette L, Traissac L, et al. EORTC Head and Neck Cancer Group. Laryngeal preservation with induction chemotherapy for hypopharyngeal squamous cell carcinoma: 10-year results of EORTC trial 24891. *Ann Oncol.* (2012) 23:2708–14. doi: 10.1093/annonc/mds065
39. Tang X, Pang T, Yan WF, Qian WL, Gong YL, Yang ZG. A novel prognostic model predicting the long-term cancer-specific survival for patients with hypopharyngeal squamous cell carcinoma. *BMC Cancer*. (2020) 20:1095. doi: 10.1186/s12885-020-07599-2
40. Heng Y, Zhu X, Zhou L, Zhang M, Li J, Tao L, et al. prognostic nomogram for predicting the long-term survival outcome of hypopharyngeal squamous cell carcinoma patients after tumour resection to assist the decision-making of postoperative adjuvant treatment. *Eur J Surg Oncol.* (2020) 46:245–51. doi: 10.1016/j.ejso.2019.09.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Liu, Tang, Zhang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fpubh.2021.812735

COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques

Zunera Jalil[1], Ahmed Abbasi[1], Abdul Rehman Javed[1], Muhammad Badruddin Khan[2], Mozaherul Hoque Abul Hasanat[2], Khalid Mahmood Malik[3] and Abdul Khader Jilani Saudagar[2]**

[1] Department of Cyber Security, Air University, Islamabad, Pakistan, [2] Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, [3] Department of Computer Science and Engineering, Oakland University Rochester, Rochester, MI, United States

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Shubham Mahajan, Shri Mata Vaishno Devi University, India
Praveen Kumar, VIT University, India
Saurabh Singh, Dongguk University Seoul, South Korea
Misbah Abbas, University of Science and Technology of China, China

***Correspondence:**
Abdul Rehman Javed
abdulrehman.cs@au.edu.pk
Abdul Khader Jilani Saudagar
aksaudagar@imamu.edu.sa

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *10 November 2021*
Accepted: *15 December 2021*
Published: *14 January 2022*

Citation:
Jalil Z, Abbasi A, Javed AR, Badruddin Khan M, Abul Hasanat MH, Malik KM and Saudagar AKJ (2022) COVID-19 Related Sentiment Analysis Using State-of-the-Art Machine Learning and Deep Learning Techniques. Front. Public Health 9:812735. doi: 10.3389/fpubh.2021.812735

The coronavirus disease 2019 (COVID-19) pandemic has influenced the everyday life of people around the globe. In general and during lockdown phases, people worldwide use social media network to state their viewpoints and general feelings concerning the pandemic that has hampered their daily lives. Twitter is one of the most commonly used social media platforms, and it showed a massive increase in tweets related to coronavirus, including positive, negative, and neutral tweets, in a minimal period. The researchers move toward the sentiment analysis and analyze the various emotions of the public toward COVID-19 due to the diverse nature of tweets. Meanwhile, people have expressed their feelings regarding the vaccinations' safety and effectiveness on social networking sites such as Twitter. As an advanced step, in this paper, our proposed approach analyzes COVID-19 by focusing on Twitter users who share their opinions on this social media networking site. The proposed approach analyzes collected tweets' sentiments for sentiment classification using various feature sets and classifiers. The early detection of COVID-19 sentiments from collected tweets allow for a better understanding and handling of the pandemic. Tweets are categorized into positive, negative, and neutral sentiment classes. We evaluate the performance of machine learning (ML) and deep learning (DL) classifiers using evaluation metrics (i.e., accuracy, precision, recall, and F1-score). Experiments prove that the proposed approach provides better accuracy of 96.66, 95.22, 94.33, and 93.88% for COVISenti, COVIDSenti_A, COVIDSenti_B, and COVIDSenti_C, respectively, compared to all other methods used in this study as well as compared to the existing approaches and traditional ML and DL algorithms.

Keywords: healthcare, COVID-19, pandemic, sentiment analysis, Twitter, internet of things

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) has severely impacted the daily lives of individuals across the globe (1). People worldwide use online media to state their viewpoints and general feelings concerning this phenomenon that has assumed control over the world by storm (2, 3). Social media platforms like Twitter have experienced exponential growth in tweets related to the pandemic in a short period (4, 5). The social networking site Twitter is a commonly used online media platform. It provides real-time information

related to ongoing events concisely and captures the emotions and thoughts of the people. During this pandemic, people use the online media platform Twitter to express their feelings, opinions, emotions, and thoughts related to the worldwide pandemic (6, 7). It rapidly spread throughout the world by an increasing number of corona cases in a short time (8). This disease has affected many countries, even the countries with hardly any or no infections because if someone is in close proximity to other people and one of them becomes affected, they will undoubtedly be impacted (9). According to Naseem et al. (1), the World Health Organization (WHO) declared COVID-19 a pandemic on January 30, 2020. All the biomedical experts worldwide are relentlessly trying to control the disease and find possible cures for this viral infection. To control this pandemic situation, vaccination is the most effective strategy to prevent the spread of COVID-19 disease all around the world (10, 11). Vaccination is the first and crucial step to stop the coronavirus outbreak. People with COVID-19 symptoms must isolate themselves and get examined, and all must get vaccinated.

During the isolation process, people express their feelings on social media; however, social media contains real-time and valuable information about COVID-19; still, data from social media might be useless or misleading at times. Sufferings get multiplied if they find misleading and depressing information on social media. With a new normal of "staying at home," "work from home," and "isolation time," social networking media has been extensively used to share news, opinions, emotions and advice. Misinformation information is defined as trying to mystify/mislead others with false or irrelevant information, such as "eating bananas is a preventative against the COVID-19 coronavirus disease." A person going through this disease passed through several physical and mental. This brings about the need to quickly apply logical strategies to comprehend informative data streams. Online Social media platforms such as Twitter and Facebook contain a lot of noisy data, so identifying informative content from large and noisy data is a challenging task, but after cleaning it, this noisy data captures human feelings and emotions, expression, and thoughts. When analyzed carefully, it conveys a lot about the present mood, attitude, and nature of a large human community.

The social media users are increasing with time because they depend on social media for informative content, and the volume of data is also increasing; this focused on the use of Natural Language Processing (NLP) with different algorithms of Artificial intelligence (AI) to extract meaningful information efficiently (12). NLP and its applications have had a significant impact on social media text analysis and classification; however, the challenges of determining a content's inherent importance using NLP-strategies, such as contextual phrases and words, ambiguity in text or speech, necessitate the use of ML-based algorithms (13–15).

In this study, we use Twitter data for sentiment analysis to identify public sentiments to investigate the increased fear associated with coronavirus. Many traditional approaches have been used to identify human behavior and nature, which presents the possibility of increasing analyses by quickly doing sentiment classification using NLP techniques. Sentiment analysis and classification of COVID-19 and other disaster-associated scenarios and keywords associated with the Twitter data analysis are essentially analyzed using the proposed methodology presented in this study outlines. This study focused on the tweets analysis and identified global people sentiments from February 2020 to March 2020. We use machine learning (ML) and deep learning (DL)-based classification methods primarily used in AI applications; however, the discussion and the comparison of tweets' sentiment classification mechanisms is one of the essential contributions of this research. Various studies in the past focused on the identification of sentiment from COVID-19 tweets but, less work is available for the identification of both the topic and sentiment considered together (16).

The tweets from February 2020 to March 2020 are collected using the Twitter API. The tweets' sentiments are categorized into three classes (positive, negative, and neutral) (1). This article focused on creating a new dataset rather than efficient categorization of users' sentiment. Therefore, we propose a new model to categorize the user's sentiments about COVID-19.

The main contributions of this study are as follows:

- Design a Transformation-based Multi-depth DistilBERT model for sentiment analysis of tweets to identify sentiments concerning coronavirus from tweets.
- Extract sentiment-related concise information from tweets to automatically learn features without human intervention.
- Present a broad comparison between existing ML and DL text classification methods and discuss the given baseline results. The proposed model outperformed on real-life datasets compared to all previously used methods.

The remainder of this article is organized as follows. Existing work related to COVID-19 sentiment analysis is presented in section 2. Section 3 provides a detailed explanation of the selected dataset. The proposed methodology is described in section 4. Experimental analysis and results are explained in section 5, and finally, the conclusion of this study is presented in section 6.

2. LITERATURE REVIEW

Since the outbreak of the coronavirus pandemic, researchers have discussed its origin, effects, and trends. This section presents the tweet sentiment analysis using different ML, DL, and NLP methods. Extracting meaningful information from noisy data is a challenging task. ML and DL techniques are necessary tools (17) to do this task. Twitter is one of the finest social media platforms for news collection (18). For sentiment analysis on India, 24,000 tweets regarding COVID-19 were crawled from Twitter (19). They just extracted the tweets related to COVID-19 and visualized the sentiments of people regarding this pandemic, and they did not perform experiments using ML techniques.

Another research focused on the topics and sentiments of people expressed on Twitter about COVID-19. They collected tweets about COVID-19 and labeled them as positive, negative, and neutral, then analyzed these tweets for sentiment classification using different feature sets and classifiers. This work used only one evaluation metric, which is accurate, and obtained the highest accuracy using the Bidirectional

TABLE 1 | Summary of the existing work.

Paper title	Methodology	Dataset	Results
COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis (34)	Proposed a COVIDSenti dataset and performed sentiment analysis using multiple classifiers	COVIDsenti	Achieved highest result using BERT with 94.80% accuracy.
Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india (19)	Visualize Sentiments on COVID-19 tweets	Crawled 24000 COVID-19 tweets	Extracts sentiments of Indian's about COVID-19
COVID-19 Public Sentiment Insights and ML for Tweets Classification (24)	Used Naive Bayes and Logistic Regression for sentiment classification	Created their dataset	Obtained 91% with Naive Bayes and 74% accuracy with logistic regression
Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach (12)	Used LSTM model for sentiment classification of COVID-19 tweets	Coronavirus Posts in Reddit Platform	LSTM model achieved an accuracy of 81.15%
Bidirectional Long Short-Term Memory Networks for Relation Classification (28)	Extract the most important features for sentiment classification using BILSTM	SemEval-2010 task 8 dataset	Achieved the best F1-score of 83.6% using BILSTM model

Encoder Representations from Transformers (BERT) model, which is 94.80%. Sometimes, we utilize only classification accuracy to assess our model's performance; nevertheless, this is insufficient to evaluate our model. We need to testify the model's performance using precision, recall, and F1-score along with accuracy (1).

Another research work focused on the psychological effect of COVID-19 to analyze the nature and prevailing mood of human behavior (20). It analyzed that people are in crisis due to coronavirus and increased anxiety levels because of COVID-19 news. Multiple studies show analysis about the industrial crisis and economic impact of the COVID-19 crisis across industries and countries (21). Over the past few years, sentiment analysis based on tweets has been utilized in numerous applications due to the large amount of data collected from various social media platforms (22). It includes Twitter, Facebook, Reddit, and YouTube. The analysis shows flaws in the collected information (23). Different ML and DL classifiers test the short and long text information. For evaluation of a short text, logistic regression, and Naive Bayes give average results of 74 and 91%, respectively, but in the case of long text testing, both the models performed very poorly (24). Recently, people have been heavily dependent on social media news, and they are conveying their viewpoints, emotion, and feelings about this novel virus *via* social media posting (25).

The recent COVID-19 studies rely on public opinion, emotion, and sentiment analysis in English social media posts from various social Critical Legal Studies (CLS). This study uses RNN models such as LSTM, BiLSTM, and SAB-LSTM to train and test the dataset. They have used the F1-score performance measure; based on this performance measure, SAB-LSTM achieved a higher F1-score when compared to LSTM, and BiLSTM Model for COVID-19 sentiment detection (26). Social media platforms such as Reddit allows healthcare service providers to collect data related to public opinions, which can be used for human behavior analysis and knowledge discovery. This study presents a generic approach based on NLP, which can extract the most critical topics about COVID-19 related comments (27). Zhang et al. (28) proposed a sentiment classification system to classify the text not at document level only at a sentence level. This approach extracts features related to public opinions and uses WordNet lexical method dataset to organize the opinion words. So this method categorized the elements under the relevant opinion sentences. The extracted features are scored according to the frequency in the reviews. The authors provide a summary based on features. They identified the most important features from the dataset and achieved the best F1-score of 83.6% using the BILSTM model. Mukherjee et al. (29) used the RNN model (bidirectional long short-term memory) to use NLP methods and utilize a bidirectional RNN to learn patterns of relations from textual data for sentiment analysis. There is much work done in tweet sentiment analysis using different DL and NLP approaches as mentioned in the study (30, 31). This study proposed a sentiment classification approach using DL and NLP models (12). NLP techniques are used for topic modeling to identify valuable topics related to coronavirus, while the LSTM model is used for classification using the sentiment-140 dataset (32). Another research gathered 1 million tweets to capture people's sentiments toward mask usage as a preventative strategy in the COVID-19 pandemic (33). They also utilized NLP to analyze the growth in the frequency of positive tweets. Our analysis finds the most common topics the public addresses in their posts concerning all preceding works. We also reported benchmarked findings for identifying tweet sentiments using NLP approaches. We compare the performance of COVID-19 tweet sentiment analysis using several ML and DL algorithms. We propose a valuable feature set with the goal of improving accuracy. The suggested system examines the sentiments of collected tweets for sentiment classification and extracts the most significant feature sets, which aids in improving classification outcomes when compared to the baseline technique. **Table 1** presents the summary of the existing work.

3. DATASET SELECTION

We used four real-world datasets for experimental analysis. The dataset is obtained through the Github repository (1) shown below to assess the classifier's classification performance. **Table 2** provides an overview of the datasets utilized in this study.

The social media users indicate more negative sentiments than the positive sentiments corresponding to the COVID-19. The trend of negative sentiments concerning time shows that people using the Twitter platform are more intended toward negative sentiments than positive sentiments during the lockdown session from February 2020 to mid of March 2020. After mid of March 2020, the negative sentiment curve dropped because people followed SOP's and kept social distancing policies applied by the Government authorities. Naseem et al. (1) divided the COVIDSenti dataset into three parts for evaluation and generalization purposes: COVIDSenti_A, COVIDSenti_B, and COVIDSenti_C, which have three sentiments, positive, negative, and neutral for classification purposes. They labeled tweets as positive, negative, and neutral by following the Fellah and Bandi guidelines (35). The COVIDSenti dataset consists of two months of tweets fetch from Twitter using Tweepy, Python Twitter API library. **Figure 1** represents the sentiment tweets in the COVIDSenti dataset.

Selection Criteria: COVIDSENTI is a two-month archive of tweets. Only English-language tweets were included in our search. The keywords guaranteed that the literary corpus focused on COVID-19 and related issues. Section 4.2 presents the keywords that are used to collect the tweets.

COVIDSenti: The COVIDSenti dataset consists of 90,000 unique tweets. These tweets are collected from 70,000 users from Twitter. A total of 2.1 million sentiment tweets from 2 months (February 2020 to March 2020) are collected. From these 90,000 unique instances, we have 6,280 positive sentiments, 16,335 negative sentiments, and 67,835 neutral sentiments for the classification task. The COVISenti dataset is further divided into three sub-datasets.

COVISenti_A: The COVIDSenti_A dataset consists of 30,000 tweets instances. From these 30,000 unique tweets, we have 1,968 positive sentiments, 5,083 negative sentiments, and 22,949 neutral sentiments for the classification task of the COVIDSent_A dataset. The COVIDSenti_A dataset contains tweets relevant to the action taken by government authorities to protect people from COVID-19. For example, "Coronavirus warning: UK threat increases as chief medical officers urge Government action."

COVIDSenti_B: The COVIDSenti_B dataset consists of 2,033 positive sentiments, 5,471 negative sentiments and 22,496 neutral sentiments with overall 30,000 unique instances. COVIDSENTI-B dataset tweets mainly relate to four topics (COVID-19 disasters, keep social distancing, lockdown, and stay at home). COVISenti_B dataset shows tweets related to lockdown and stay at home even if one suffers from the disease. For example, "Stay Home If You Are Sick, or If you are sick stay home regardless of what you have."

COVIDSenti_C: This dataset contains 2,279 positive sentiments, 5,781 negative sentiments, and 21,940 neutral sentiments, collectively 30,000 tweets from the Twitter platform related to three different topics (COVID-19 cases, stay at home, and outbreak). The COVIDSenti_C is a collection of tweets on COVID-19 cases, outbreaks, and stay-at-home advice. We find people's behavior from these topics and analyze that COVID-19 cases are increasing daily. For example, "Airport screenings for the Wuhan coronavirus increase around the world."

TABLE 2 | Overview of datasets.

Dataset\Label	Positive	Negative	Neutral	Total
COVIDSenti-A	1,968	5,083	22,949	30,000
COVIDSenti-B	2,033	5,471	22,496	30,000
COVIDSenti-C	2,279	5,781	21,940	30,000
COVIDSenti	6,280	16,335	67,835	90,000

4. PROPOSED APPROACH

The proposed approach is divided into four phases: 1) pre-processing, 2) keyword trend analysis, 3) word embeddings for feature extraction, and 4) classification methods. The CovidSenti

FIGURE 1 | Representation of sentiment tweets in COVIDSenti.

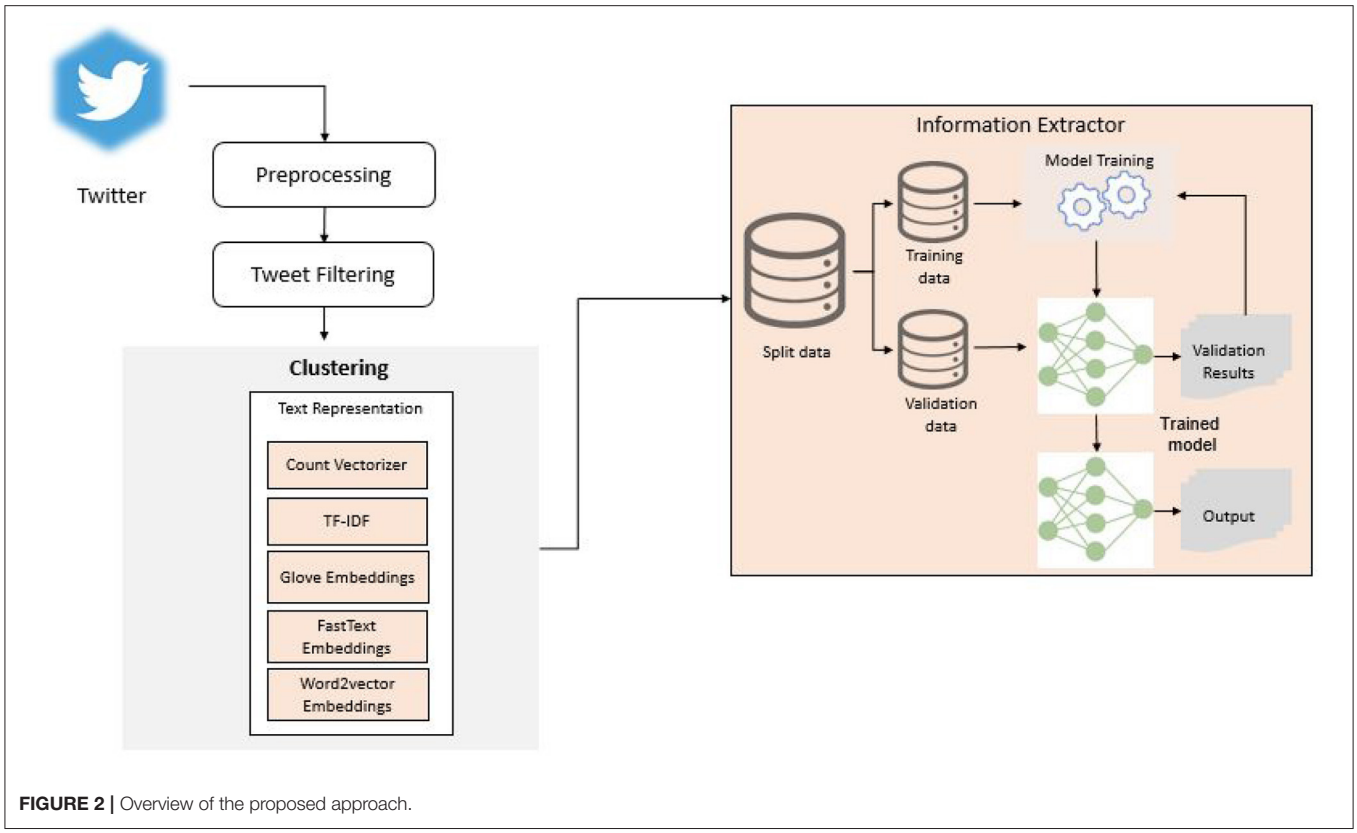


FIGURE 2 | Overview of the proposed approach.

dataset is divided into two chunks, training and testing. We take care of the various factors of the dataset, such as over-fitting, noisy or small and large datasets. The main objective of this study is to evaluate the classification performance of state-of-the-art classifiers on the COVIDSenti dataset and then attempt to improve performance by extracting key features of tweets. The proposed technique classifies the CovidSenti dataset with higher accuracy and competently for the COVIDSenti dataset containing COVID-19 associated Twitter posts. **Figure 2** demonstrates our proposed approach with each of the methods explained in the following figure.

4.1. Data Pre-processing

Information gathered from social networking media platforms is more often noisy and heterogeneous. We make the Twitter stream ready for exploratory analysis; the pre-processing step first changes the uppercase letters to lower case, then removes all the special characters, stop words, mentions, and URLs from the dataset tweets. For hashtag, first, it separates the total hashtags into tiny fragments as divided hashtags that positively affect the data clusters. A few hashtags are written using camel case, for example, "#StayHome," which are easy to convert into the segment. Still, on the other hand, some hashtags that do not involve any camel case, e.g., "#stayhome," a vast vocabulary is needed to discover the longest string suits in the hashtag. The pre-processing step uses a set of vocabulary of almost 70,000 English words to handle these challenges. Identifying informative content from a large and noisy dataset such as tweets is a challenging task. To achieve this, the following techniques are carried out inside the given order to enhance the text.

1. One common way to analyze COVIDSenti data is to calculate word frequencies to understand how often words are used in tweets. So, the first step is lemmatization that processes with the use of a vocabulary and morphological analysis of phrases and returns root words. We used lemmatization with the nltk method that converts a phrase to its base form, for example, "deaths" to "death" or "caring" to "care").
2. The second step is to remove the stop words. It is the most suitable technique to overcome the noise from the textual tweets (such as "the," "a," "an," "in"). Stop words can be filtered from the text to be processed, and it does no longer affect understanding of a tweet sentence's valence. We removed the most common stop-words that are present in a text, for example, "a," "an," "in," and "the."
3. Some time model understand the actual word as two different words because words are taken as case sensitive (i.e., COVID, covid). To avoid it, we convert all capital letters or words into the lower case. This method does not change the word's actual meaning or original word.
4. Contraction is a process of shortening the by through replacing or dropping letters with the aid of an apostrophe. Nowadays, people move to online media to connect with others *via* textual content or posts like Facebook, WhatsApp, Instagram, and Twitter. Many people communicate with each

other; people mostly use shortened forms and abbreviations of words in their text. We used the contraction mapping method that drops the vowels from the words. Removal of contraction mapping is related to text standardization, and it is helpful while working with Twitter data in sentiment analysis.
The main advantage of using textblob is its many capabilities like noun phrase extraction, pos-tagging, and sentiment analysis.

5. Part of speech is the essential step of pre-processing. It is essential to build parse trees, which might be utilized in constructing "most named entities are nouns" (NERS). It is also used to extract relations among words. Part of speech tagging is essential for building lemmatizers to return a phrase to its root form.
6. Twitter data is noisy, which affects the performance of a classifier, so the pre-processor eliminates URLs, @user_mentions. We remove alphanumeric or special characters and remove non-ASCII characters and numbers from our dataset because they do not help us detect sentiment. We also replace emojis with their corresponding reaction in text.
7. For hashtags, we eliminate the "#" symbol from the start of the phrase. We used tokenizer to split hashtags into appropriate words, for example, "#stayhomestaysafe," tokenizer converted it into "stay," "home," "stay," "safe." Many words are concatenated in other words, and we also performed word-segmentation to achieve this.

4.2. Keyword Trend Analysis

We performed keyword trend analysis on our pre-processed corpus to find the most frequent words. The top 10 most commonly used keywords are "Coronavirus", "Corona", "COVID-19", "Virus", "Coronavirus Cases", "Cases", "Social distance", "Coronavirus pandemic", "Coronavirus Crisis", and "Stay home" with 156,668, 87,661, 26,239, 22,289, 16,638, 10,269, 5,768, 3,125, 1,981 and 1,181 number of frequencies, respectively. We conclude that most people talk about coronavirus cases, the social distancing, COVID-19 outbreak, the coronavirus crisis, and stay-home.

To analyze the main topics across the dataset, we discover the topic distributions using Glove embeddings. We used Glove embeddings because it has an extensive vocabulary size. We can also find the words from COVIDSenti data that are not present in the Glove embeddings, such as contractions, misspelled words, concerted words, or emojis, which can decrease our model's performance. We discovered that most social media users discuss coronavirus cases, the coronavirus epidemic, social distance, the need to stay at home, the coronavirus crisis, and the crises caused by the coronavirus. Based on this analysis, we examine the distribution of the 15 topics across the corpora, in which topics 1 ("coronavirus"), 2 ("virus"), and 3 ("corona") were the top three in the entire corpora.

4.3. Feature Extraction

In this study, count vectorizer, TF-IDF, and word embeddings techniques are used for feature extraction. The count vectorizer feature extraction technique is used to convert given tweets into a vector space, and it covers the tweets based on the most commonly used words (count) that frequently occur in the tweets. The count vectorizer creates a word matrix where every unique word represents the column of the matrix, and the selected text from the document represents the row of the matrix. In this way, we count the word in that particular text sample. We also used the term frequency-inverse document frequency (TF-IDF) feature extraction technique along with a count vectorizer. In this study, the TF-IDF is used for tweet analysis, and it has weighted features for execution boosting. The TF-IDF takes the TF and its corresponding IDF as a product to get the weights of features in a document. The length determines the TF of features in a single document. It is defined in Equation (1).

$$TF = \frac{count_{t,d}}{totalcount_d} \tag{1}$$

In equation 1, the number of TF t in document d is represented by $count_{t,d}$, while the overall number of terms in that document is represented by $totalcount_d$. IDF thinks that the text's increase in term t will be more informative for model training. It can be defined as in equation 2.

$$idf^{'} = i^{'}/df_t^{'} \tag{2}$$

Where $i^{'}$ is the total number of documents, and $df_t^{'}$ is the number of documents that include the phrase t. When a term t frequently appears in many documents, IDF computes the weights of a phrase t low. For example, stop words have a lower IDF value. So finally, the TF-IDF can be defined as in equation 3.

$$tf - idf = tf_{t,d} \quad * \quad \log(idf). \tag{3}$$

We also employed the word embeddings model for feature extraction, such as pre-trained models Word2Vec, Glove, and fastText embeddings with 300-D vectors. Furthermore, we used RCNN hybrid models for better sentiment classification tasks and reduced the over-fitting by decreasing bias because there are many variances. We also used transformer-based language models that have been widely used in the NLP research area. We used Multi-depth DistilBERT transformer-based model to pre-train a smaller general-purpose language representation model. We fine-tuned the Multi-depth DistilBERT model to achieve high performance on multiple tasks like better sentiment classification. Most of the previous work investigated that the used-ability of the distillation is only for building task-specific models. Still, the primary purpose is to decrease the size of the BERT transformer model. We decreased the BERT transformer model by 40% and retained its actual language understanding capabilities by 97%, and it was 60% faster than BERT.

These transformer-based language models (LMS) show how the training methods are used. We used unsupervised learning in the training process to generate an LM. The NLP models are trained on extensive data; these models attain more improved context word representations than conventional and non-contextual word representation methods because of the large data size.

4.4. Classification Methods

To analyze the classification performance on the COVIDSenti dataset, we used different ML and DL-based classifiers to measure the performance in the sentiment classification tasks.

We used XGBoost (eXtreme Gradient Boosting) in our analysis. XGB classifier is similar to the Gradient Boosting classifier. It consists of multiple trees, and it is a tree base model which is why it has gained lots of attention in the past few years. Many weak learners are working parallelly distinct to Gradient Boosting; XGboost gives a speed boost because of this technique. XGboost uses L1 and L2 regularization methods to control over-fitting, which are not present in Adaboost and Gradient Boosting techniques. XGBoost has a new feature of scalability so that it performs better in the distributed environment than a single system. At every iteration of XGBoost, we calculate some errors. We used this error to correct the previous prediction and optimize the loss function. Regularizer is utilized in the loss function to evaluate the classifier performance, which is defined by

$$X^{'}(\otimes) = L^{'}(\otimes) + \Omega(\otimes). \quad (4)$$

In equation (4), we train the parameters using $\otimes$, for the training Loss function, we used $L^{'}$, and to measure the model's complexity, we used ω regularization. In this study, we set different parameters for the proposed XGBoost classifier. The XGB classifier is composed of 1,000 n_estimators with a 0.1 learning rate. Learning rate (LR) is used to handle the over-fitting. The XGB used a minimum of one child weight and a maximum depth of 6.

For DL-based models, we used the Conv1D-LSTM model. While using the DL classifier, we used Adam optimizer and L2 regularization. We trained the model on 50 epochs. Our Conv1D-LSTM consisted of four layers: the embedding layer with weights and vocab_size as input, then the Conv1D layer with a relu activation function. Next, we used the LSTM layer, and then at the end, the dense layer is used, also called an output layer. The softmax is employed as an activation function. In the experiments, we used a hybrid model named RCNN. We used this hybrid model for the sentiment classification task. RCNN model consisted of six layers in which one CNN layer with relu activation function, and then we used BiLSTM, an LSTM layer. After that, we used a dropout of 0.4 to control over-fitting; finally, we used a dense layer with softmax as an activation function. The RCNN uses Adam optimizer with verbose = None. We used the Multi-depth DistilBERT transformation model at the end of our experiments. The parameters of the Multi-depth DistilBERT model are presented in **Table 3**.

TABLE 3 | The parameter settings of multi-depth DistilBERT.

Value	Setting
Model name	distilbert-base-uncased
Number of epochs	3
Batch size	80
Max sequence length	256
Learning rate	5e-05
Accumulation steps	4
Random seed	42

5. EXPERIMENTAL ANALYSIS AND RESULTS

The main aim of this study is to evaluate the classification performance on COVIDSenti datasets and provide benchmarked results. This paper applies ML, DL, and hybrid methods to COVIDSenti datasets and examines their performance using several evaluation measures. The performance assessment metrics in this study include accuracy, precision, recall, and F1-score. These standard performance indicators have been carefully chosen to attest to the model's capacity to generate the best categorization performance. Following the experimentation procedure, the experimental outcomes are compared to state-of-art-methodologies.

5.1. Results

This work consists of different ML, DL, hybrid-based, and transformer-based models on four COVIDSenti datasets and analyzes their performance using the defined evaluation metrics. The proposed approach is used to gauge the performance in the sentiment classification task. The highlighted results in **Tables 4–9** represent the highest achieved accuracy as compared to baseline results.

5.1.1. COVIDSenti

Figure 3 and **Tables 4–9** present the results achieved on the COVIDSenti dataset for tweet sentiment classification. In the case of the count vectorizer feature extraction technique, for sentiment classification on the COVIDSenti dataset, the XGB model attains the best accuracy of 89.81%, other ML techniques, such as KNN, LR, and Ensemble model, achieve the accuracy of 79.21, 89.03, and 88.75%, respectively. The xgb model achieves the highest accuracy of 89.81% with 89% precision, 90% recall, and 89% F1-score compared to KNN, LR, and Ensemble model.

While using the TF-IDF feature extraction technique for sentiment classification again, the xgb model outperforms other ML models, such as SVM, RF, NB, and DT with the highest accuracy of 88.46% with 88% precision, 88% recall, and 88% F1-score. We also used different word embedding techniques, such as FastText, Word2Vec, and Glove with different ML classifiers. We fine-tuned FastText with a Glove, resulting in 80% accuracy on the COVIDSenti dataset. We also measure other evaluation metrics like precision, recall, F1-score, 79% precision for the COVIDSenti dataset, 79% recall, and 73% F1-score. While working with the Word2Vec word embedding technique, we used the xgb model and achieved an accuracy of 79.71% which is the highest accuracy compared to other Word2Vec models (RF with 79.09% and DT with 76.06%). Along with the accuracy, we measure the precision, recall, and F1-score of the xgb model, which is 79, 79, and 74%, respectively. Finally, we employ the Glove word embedding technique with the ML model. We fine-tuned Glove with the xgb model and gained an accuracy of

TABLE 4 | Machine learning (ML) classifiers accuracy using count vectorizers.

Model dataset		COVIDSenti	COVIDSenti_A	COVIDSenti_B	COVIDSenti_C
		Proposed model/Accuracy			
Count vectorizer	KNN	79.21%	79.65%	77.83%	77.75%
	LR	90.03%	87.75%	86.76%	86.04%
	Ensemble	88.75%	88.51%	87.08%	86.36%
	XGB	**89.81%**	**88.71%**	**88.03%**	**87.07%**

The bold ones represent better results.

TABLE 5 | Comparison of proposed deep learning (DL) Classifiers accuracy with baseline using word embeddings.

Model dataset		COVIDSenti	COVIDSenti_A	COVIDSenti_B	COVIDSenti_C
Existing models/Accuracy					
Word2vec	BiLSTM	76.09%%	76.05%	74.05%	74.09%
Glove	BiLstm	77.01%	76.08%	72.09%	74.03%
DCNN- (Glove+ CNN)		86.09%	83.04%	83.02%	86.04%
Proposed model/Accuracy					
Conv1D-LSTM + Glove		**87.06%**	**86.10%**	**84.44%**	**86.90%**
Gain		**0.97%**	**3.06%**	**1.42%**	**0.86%**

The bold ones represent better results.

TABLE 6 | Comparison of proposed hybrid model accuracy with baseline.

Model dataset	COVIDSENTI	Covidsenti_A	Covidsenti_B	CovidSenti_C
Existing models/Accuracy				
IWV	77.01%	76.03%	74.09%	73.05%
HyRank	88.01%	85.04%	86.05%	87.07%
Proposed model/Accuracy				
RCNN	**95.40%**	**92.90%**	**92.93%**	**93.27%**
Gain	**7.39%**	**7.86%**	**6.88%**	**6.20%**

The bold ones represent better results.

TABLE 7 | Comparison of proposed ML classifiers accuracy with baseline using term frequency-inverse document frequency (TF-IDF).

Model dataset		COVIDSenti	COVIDSenti_A	COVIDSenti_B	COVIDSenti_C
Existing models/Accuracy					
TF-IDF	SVM	84.05%	83.09%	83.00%	82.08%
	RF	84.01%	83.06%	82.08%	82.01%
	NB	77.03%	76.05%	75.00%	73.01%
	DT	79.04%	78.01%	76.08%	75.03%
Proposed model/Accuracy					
	XGB	**88.46%**	**88.31%**	**87.41%**	**86.03%**
	Gain	**4.41%**	**5.22%**	**4.41%**	**3.95%**

The bold ones represent better results.

79.19% with the precision of 79%, recall of 79%, and F1-score of 74%. After ML with word embeddings, we employ DL with word embeddings. The proposed model Conv1D-LSTM with Glove gains an accuracy of 87.60%. The best precision, recall, and F1-score were gained from the Conv1D-LSTM classifier, 87, 87, and 88%, respectively.

We used the RCNN hybrid model for the sentiment classification task, and it performs very well on the COVIDSenti dataset. The best accuracy of hybrid models is achieved from the RCNN model, which is 95.40%. RCNN model shows a better 95, 96, and 95% in terms of precision, recall, and F1-score. Finally, we used a transformer-based language classifier which shows the overall highest accuracy of 96.66% with the highest precision of 96% also with 96% recall and 95% accuracy.

5.1.2. COVIDSenti_A

The results achieved on the COVIDSenti_A dataset are presented in **Tables 4–9** and plotted in **Figure 3**.

We used the count vectorizer feature extraction technique to classify sentiments first. The highlighted XGB model gains the highest accuracy of 88.71% compared to other ML models: KNN, LR, and Ensemble, 79.65, 87.75, and 88.51%, respectively. We also measure the other evaluation metrics (precision, recall, and F1-score) for the XGB classifier. The precision score of the XGB model for the COVIDSenti_A dataset is 88%, recall score is 89%, and the F1-score score is 88%. We also used different ML classifiers with the TF-IDF feature extraction technique. The XGB classifier on the COVIDSenti_A dataset exhibits the highest accuracy of 88.31% with 88% precision, 88% recall, and also 88% F1-score as compared to other ML models such as SVM, RF,

TABLE 8 | Comparison of proposed ML classifiers accuracy with baseline using word embeddings.

Model dataset		COVIDSENTI	Covidsenti_A	Covidsenti_B	CovidSenti_C
Existing models/Accuracy					
Word2Vec	RF	76.09%	76.04%	73.02%	75.03%
	DT	76.06%	76.02%	74.05%	74.01%
Glove	RF	72.06%	71.06%	70.02%	72.02%
	DT	70.01%	69.03%	68.09%	69.04%
FastText	SVM	81.05%	80.01%	79.02%	78.03%
	NB	73.05%	73.02%	74.05%	72.01%
	RF	84.05%	82.03%	84.01%	80.02%
Proposed model/Accuracy					
Fine tuned FastText	XGB	80.00%	78.33%	77.61%	76.30%
Fine tuned Word2Vec	XGB	79.71%	78.33%	77.85%	76.05%
Fine tuned Glove	**XGB**	**79.19%**	**78.39%**	**78.23%**	**76.05%**
Loss		**4.05%**	**3.64%**	**5.78%**	**3.97%**

The bold ones represent better results.

TABLE 9 | Comparison of proposed transformer based model accuracy with baseline models.

Model dataset	COVIDSenti	COVIDSenti_A	COVIDSenti_B	COVIDSenti_C
Existing models/Accuracy				
distilBERT	93.09%	93.07%	92.09%	92.06%
BERT	94.08%	94.01%	93.07%	92.02%
XLNET	93.03%	92.04%	91.04%	92.00%
ALBERT	92.09%	91.04%	92.00%	91.01%
Proposed model/Accuracy				
Multi-depth DistilBERT	**96.66%**	**95.22%**	**94.33%**	**93.88%**
Gain	**2.58%**	**1.21%**	**1.26%**	**1.82%**

The bold ones represent better results.

NB, and DT. We also apply different word embedding techniques (FastText, Word2Vec, and Glove) on the COVIDSenti_A dataset.

In the case of FastText word embedding with XGB classifier on COVIDSenti_A, we achieve 78.33% accuracy with 77% precision, 79% recall, and 73% F1-score. We also used Word2Vec embedding with XGB classifier on COVIDSenti_A dataset. XGB classifier with Word2vec word embedding achieves an accuracy of 78.33% which is the highest accuracy compared to other word2vec techniques with ML classifiers. The XGB classifier with Word2Vec gains a precision score of 77%, recall score of 78%, and F1-score of 73%. Finally, we used Fine-tuned Glove word embedding with XGB classifier, which achieves the highest accuracy compared to Word2vec and Glove word embeddings with ML. The fine-tuned Glove with XGB classifier achieves an accuracy of 78.39% on the COVIDSenti_A dataset. We also measure other performance metrics to testify to the model's capability. Other performance metrics are precision with 77%, recall with 79%, and F1-score with 73%.

After ML word embedding, we move on to DL word embedding for better sentiment classification. While working with DL classifiers with word embeddings, we apply Conv1D-LSTM with the Glove word embedding technique. Conv1D-LSTM with Glove word embedding achieves the best accuracy of 86.10% with other performance metrics such as precision, recall, and F1-score. The Conv1D-LSTM model observes the maximum F1-Score, precision, and recall, i.e., 86%, 86%, 87%, respectively, while having a maximum accuracy of 86.10%. We propose a hybrid model named RCNN. The best performance is achieved on the COVIDSenti_A dataset in terms of accuracy, precision, recall, and F1-score by the RCNN model with a maximum accuracy of 92.90% and maximum precision-recall, and F1-score of 92, 92, and 92%, respectively, as compared to other hybrid models. Finally, we proposed a Multi-depth DistilBERT transformer-based model, which performed very well on the COVIDSenti_A dataset and all other methods on the COVIDSenti_A dataset. The maximum accuracy on COVIDSenti_A using Multi-depth DistilBERT is 95.22% with precision, recall, and F1-score of 95%, 96%, 95%, respectively.

5.1.3. COVIDSenti_B

Figure 3 and **Tables 4**–**9** depict the sentiment classificaition results in the COVIDSenti_B dataset by different ML, DL, Hybrid models, and transformer based models. First, we used the count vectorizer feature extraction technique to extract useful features and pass them to models for sentiment classification tasks. We used the XGB model for sentiment classification on the COVIDSenti_B dataset. The XGB model obtains the maximum accuracy of 88.03% compared to other ML models such as KNN, LR, and Ensemble, which acquire 79.65, 87.75, and 88.51% accuracy. The best precision, recall, and F1-score on COVIDSenti_B are gained from the XGB classifier, 88, 89, and 88%, respectively.

Another feature extraction technique is used with ML classifiers for sentiment classification. Features are extracted

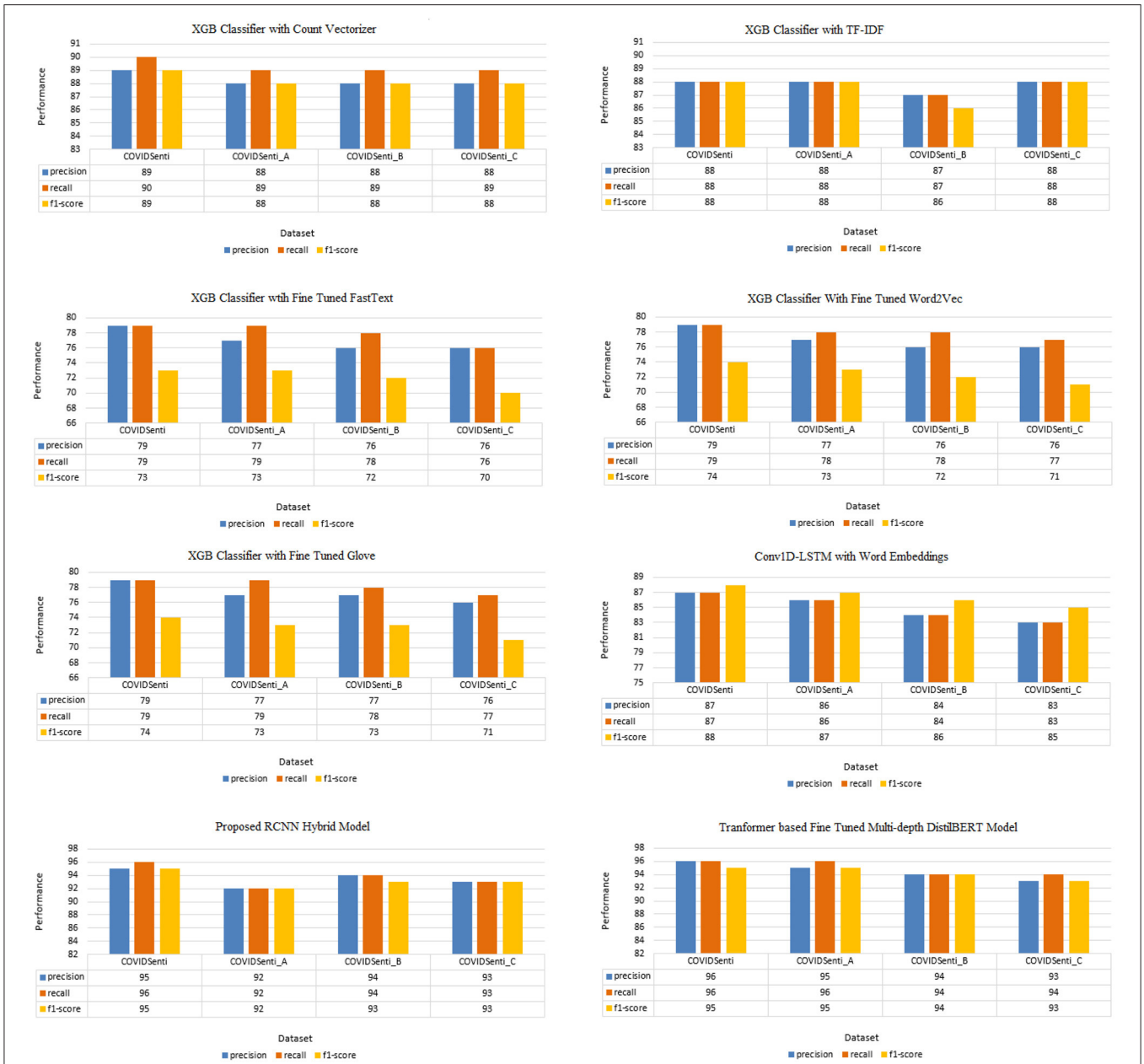


FIGURE 3 | Precision, Recall, and F1-score of classifiers used in proposed approach (percentage results).

using TF-IDF, and then we apply ML classifiers on these features for COVIDSenti_B dataset sentiment classification. We pass these TF-IDF features to the XGB model for better classification results, and XGB shows the accuracy of 87.41% on the COVIDSenti_B dataset with better precision, recall, and F1-score of 87, 87, and 86%, respectively. We used three different word embedding techniques with ML models to reduce the computational cost. First, we used Fine-tuned Glove with XGB ML model, which obtains an accuracy of 78.23% which is quite well as compared to other Word2Vec and Glove word embedding in **Table 8**. With the best accuracy of 78.23%, the Fine-tuned Glove with the XGB ML model gets the precision, recall, and F1-score of 77, 78, and 73%, respectively.

Second, we used the Word2Vec word embedding technique using the XGB ML model. The fine-tuned Word2Vec with XGB model gets the accuracy of 77.85% on the COVIDSenti_B dataset with a precision score of 76%, recall 78%, and F1-score of 72%. In the end, we used fine-tuned FastText word embedding with an XGB ML classifier. The fine-tuned FastText word embedding with XGB ML classifier gains less accuracy of 77.61% compared to proposed Glove and Word2Vec embeddings methods. The precision, recall, and F1-score of fine-tuned

FastText word embedding with XGB ML classifier is 76, 78, and 72%, respectively.

We used Conv1D-LSTM with Glove word embedding for better results, so we used DL with word embeddings to achieve the highest results for sentiment classification. For this purpose, we train the Conv1D-LSTM model on the COVIDSenti_B dataset and get the evaluation result of 84.44% with other different performance metrics. The other performance metrics are precision, recall, and F1-score with 84, 84, and 86%, respectively. We also proposed a hybrid model by adding multiple layers of LSTM, BiLSTM, and CNN for better classification on the COVIDSenti_B dataset. To achieve better classification on the COVIDSenti_B dataset, RCNN hybrid model gets 92.93% with better precision, recall, and F1-score. The precision, recall, and F1-score of the RCNN model are 94, 94, and 93%, respectively. Finally, we apply a transformer-based model named Multi-depth DistilBERT on the COVIDSenti_B dataset, which gives an accuracy of 94.33% with 94% precision, 94% recall, and 94% F1-score. F1-score of 94% shows that this model performed very well on the COVIDSenti_B dataset.

5.1.4. COVIDSenti_C

The results of the COVIDSenti_C dataset are shown in **Tables 4**–**9** and plotted in **Figure 3**. We used different ML, DL, hybrid models, and transformer-based models for COVIDSenti_C dataset classification. We used the count vectorizer technique to extract essential features from the COVIDSenti_C dataset as shown in **Table 4**. When we extract essential features, we send these features to ML models to perform sentiment classification tasks. We used KNN, LR, Ensemble, and XGB models for sentiment classification on the COVIDSenti_C dataset. For sentiment classification in the COVIDSenti_C dataset, the XGB model obtains the best accuracy of 87.07%. Other conventional techniques, KNN classifier, LR classifier, and Ensemble classifier, get the accuracy of 77.75, 86.04, and 86.36%, respectively. XGB classifier shows a proficient gain of 88, 89, and 88% in terms of precision, recall, and F1-score, respectively, compared to other models.

Furthermore, we apply another technique to extract important features for better results. We extract important features using TF-IDF and send those features to multiple ML classifiers for classification purposes. Finally, the XGB classifier using these TF-IDF features delivers the best accuracy of 88.31% on the COVIDSenti_C dataset. The precision, recall, and F1-score of the XGB model with 88.31% accuracy are 88, 88, and 88%, as shown in **Figure 3**. The computational time plays an important role during the model's training process, so to reduce computational time, we apply different word embeddings to ML and DL models. For ML, first, we apply FastText with the XGB model. We fine-tuned FastText word embedding and passed this word embedding to the XGB model to reduce computational time, which results in 76.30% accuracy, 76% precision, 76% recall, and 70% F1-score. Less F1-score indicates that we need to improve the classification result. The second embedding technique to reduce computational time is Word2Vec. While using fine-tuned Word2Vec embeddings with the XGB model, we achieved better accuracy, 76.05%, and quite well precision, recall, and F1-score, which is 76, 77, and 71%. The last and essential word embedding technique we used in this study is Glove. While using fine-tuned Glove with the XGB classifier again, we obtained an accuracy of 76.05% with the same precision, recall, and F1-score of 76, 77, and 71%. We moved to DL with word embeddings; first, we used Conv1D with LSTM and Glove. This proposed model worked very well on the COVIDSenti_C dataset and gave the accuracy of 86.91% which is the highest accuracy of the COVIDSenti_C dataset as shown in **Table 5** and also gives the best precision, recall, and F1-score of 83, 83, and 85% shown in **Figure 3**.

Furthermore, we propose our hybrid model, which performed very well on the COVIDSenti_C dataset. The proposed RCNN model achieves the best accuracy of 93.27% which is the highest accuracy in the COVIDSenti_C dataset is shown in **Table 6**. The evaluation metrics include precision, recall, and F1-score with 93% precision, 93% recall, and 93% F1-score. To cover all the methods with better classification results, we proposed a transformer-based language model that achieves the highest accuracy compared to all other methods used with COVIDSenti_C. The Multi-depth DistilBERT gives the highest accuracy of 93.88%, which is the highest accuracy of the COVIDSenti_C dataset compared to all other methods applied on the COVIDSenti_C dataset. It also gives the highest precision, recall, and F1-score of 93%, 94%, and 93% on the COVIDSent_C dataset.

5.2. Comparative Analysis With Baseline Approach

To assess the performance of the suggested strategy, we compare our experimental findings to those of the state-of-the-art approach (1), whose experimental circumstances are identical to those used in our work. The baseline approach used only accuracy as an evaluation metric to testify classifier's ability. As compared to baseline proposed approach used precision, recall, and F1-score along with accuracy metric. The proposed framework analyzes tweets' sentiments for sentiment classification and extracts the most important features that help increase the classification results compared to the baseline approach. Furthermore, we proposed a fine-tuned transformer-based Multi-depth DistilBERT model that achieved the highest results in terms of accuracy as compared to all the baseline results. The comparative results are reported in **Tables 4**–**9**. To achieve the higher performance for ML classifiers, we used different conventional methods: count-vectorizer, TF-IDF, different word embeddings based models like Word2Vec, fastText, and Glove. The results of count-vectorizer-based classification are shown in **Table 4**.

When using count vectorizer with KNN, logistic regression, ensemble method, and XGB ML classifiers on COVIDSenti datasets, the count vectorizer with XGB classifier exhibited better performance on the COVIDSENTI dataset than other ML classifiers(KNN, LR, and Ensemble). The classification accuracy of the XGB model using the count vectorizer method is 89.81% on COVIDSenti, 88.71% on COVIDSenit_A, 88.03% on COVIDSenit_B, and 87.07% on COVIDSenti_C, respectively.

In addition, we conducted a comparison using TF-IDF-based classification. The obtained results are shown in **Table 7**. On all COVIDSenti datasets, the TF-IDF was used in conjunction with the SVM, RF, NB, DT, and XGB models. Compared to other baseline ML classifiers, the TF-IDF approach with XGB classifier performs better on the COVIDSENTI-A dataset, COVIDSENTI-B dataset, COVIDSENTI-C dataset, and COVIDSENTI dataset. The classification accuracy of the XGB model using the TF-IDF method is 88.46% with 4.41% gain on COVIDSenti, 88.31% accuracy with 5.22% gain on COVIDSenit_A, 87.41% accuracy with 4.41% gain on COVIDSenit_B, and 86.03% accuracy with 3.95% gain on COVIDSenti_C, respectively. For the comparison of ML with word embeddings, we employed different word embeddings techniques like FastText, Glove, and Word2Vec with various ML classifiers: RF, DT, SVM, NB, and XGB as shown in **Table 8**. In the case of word2vec embedding, we fine-tuned word2vec with an XGB classifier, which shows better performance (79.17, 78.33, 77.85, and 76.07% on COVIDSENTI, COVIDSENTI-A, COVIDSENTI-B, and COVIDSENTI-C, respectively) as compared to baseline ML models with word2vec embeddings.

Furthermore, we fine-tuned the Glove with an xgb classifier for Glove word embedding, which outperforms the baseline Glove embedding with different ML classifiers. The xgb model achieves the highest accuracy of 79.19% on COVIDSenti, 78.39% on COVIDSenti_A, 78.23% on COVIDSenti_B, and 76.05% on COVIDSentiC as compared to baseline ML models. We detect a decrease in accuracy for the proposed FastText embedding model that is minor when compared to the yield inaccuracy for other word embedding techniques. We get the loss in terms of accuracy while comparing the proposed DL classifier with the word embedding approach to the baseline FastText approach, which is 4.05% for COVIDSenti, 3.64% for COVIDSenti_A, 5.78% for COVIDSenti_B, and 3.97% for COVIDSenti_C datasets. To compare DL-based classifiers, we apply various word embeddings models like Glove and 1D convolutional neural network (1DCNN) with LSTM, where Glove is utilized for word representations. The comparative results of DL-based classifiers are presented in **Table 5** with baseline, Word2Vec, and Glove with DL classifiers on all three COVIDSenti datasets. Compared to baseline word embedding-based classifiers, our proposed approach Conv1D-LSTM with Glove showed better performance (87.60% with 0.97% gain, 86.10% with 3.06% gain, 86.44% with 1.42%, and 86.90% with 0.86% accuracy gain on COVIDSenti, COVIDSenti_A, COVIDSenti_B, and COVIDSenti_C).

In **Table 6**, we also compare hybrid models to baseline models such as the hybrid ranking model and IWV model. The baseline approach used a hybrid ranking model that includes sentiments and context of Twitter posts for Twitter sentiment analysis (36). The comparative results with baseline Hybrid ranking and IWV model are shown in **Table 6**. Notice that the proposed hybrid model (RCNN) outperforms the baseline Hybrid ranking model and IWV model with the performance score of 95.40% with 7.39% accuracy gain, 92.90% with 7.86% accuracy gain, 92.93% with 6.88% accuracy gain, and 93.27% with 6.20% accuracy gain on COVIDSenit, COVIDSenti_A, COVIDSenti_B, and COVIDSenti_C, respectively.

Table 9 shows the final comparison of transformer-based models. The previous study fine tuned the transformer-based models such as BERT, DistilBERT, XLNET, and ALBERT. Compared to previously fine-tuned transformer-based models, we fine-tuned the Multi-depth DistilBERT model that performed very well compared to all other transformer-based models and all the other approaches used before in this study. Experiments prove that the proposed Multi-depth DistilBERT model achieves better accuracy of 96.66% with 2.58% gain on COVIDSenti, 95.22% with 1.21% gain on COVIDSenti_A, 94.33% with 1.26% gain on COVIDSenti_B, and 93.88% with 1.82% gain on COVIDSenti_C, respectively.

6. CONCLUSION

Since the outbreak of the COVID-19 pandemic and with a new normal of staying at home, working from home, and "isolation time," social networking media has been extensively used to share news, opinions, emotions, advice; however, most of the data on social media are irrelevant and do not belong to the actual scenario. This study proposed an approach to deal with the Twitter sentiment using the COVIDSenti dataset. We evaluate ML and DL classifiers using novel feature extracting methods that automatically learn features without human interference. We observed that people follow government policies and Standard Operating Procedures (SOPs) and began to favor lockdown and keep social distancing in March 2020, but the order by the government is in February 2020. There is much misinformation on social media; therefore, health organizations need to develop a stable system for detecting coronavirus precisely to preclude the spread of fake news. The proposed approach performed very well on the given dataset and showed higher accuracy when compared to similar state-of-the-art studies. In future work, we plan to analyze public sentiments toward other essential topics, such as government response to the pandemic situation, healthcare facilities by government, offline examination, and mental health by using DL algorithms to increase their performance on the dataset. One limitation of this work is that it is specific and does not look at the mood and emotions of the people. Further work can be done on the detection of mood-based sentiment analysis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

AUTHOR CONTRIBUTIONS

ZJ, AJ, and AA: conceptualization. AA: data curation. AJ: formal analysis, investigation, methodology, and software. AS: funding

acquisition. KM, AS, MB, and MA: project administration. ZJ, MB, and MA: resources. KM, AS, AJ, KM, and AS: supervision. AA and ZJ: validation. AA, KM, and AS: visualization. MB, MA, and ZJ: writing–review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, for funding this research work through project number 959.

REFERENCES

1. Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Trans Comput Soc Syst*. (2021) 8:1003–15. doi: 10.1109/TCSS.2021.3051189
2. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst*. (2021) 1–15. doi: 10.1007/s00530-021-00774-w
3. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc*. (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
4. Mansoor M, Gurumurthy K, Prasad V. Global Sentiment Analysis Of COVID-19 Tweets Over Time. *arXiv preprint* arXiv:201014234. (2020).
5. Mittal M, Siriaraya P, Lee C, Kawai Y, Yoshikawa T, Shimojo S. Accurate spatial mapping of social media data with physical locations. In: *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA: IEEE (2019). p. 4113–6.
6. Chauhan S, Banerjee R, Chakraborty C, Mittal M, Shiva A, Ravi V. A self-congruence and impulse buying effect on user's shopping behaviour over social networking sites: an empirical study. *Int J Pervasive Comput Commun*. (2021) doi: 10.1108/IJPCC-01-2021-0013
7. Chauhan S, Mittal M, Woźniak M, Gupta S, Pérez de Prado R. A technology acceptance model-based analytics for online mobile games using machine learning techniques. *Symmetry*. (2021) 13:1545. doi: 10.3390/sym13081545
8. Dhama K, Khan S, Tiwari R, Sircar S, Bhat S, Malik YS, et al. Coronavirus disease 2019-COVID-19. *Clin Microbiol Rev*. (2020) 33:e00028–20. doi: 10.1128/CMR.00028-20
9. Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in pakistan. *Comput Mater Continua*. (2021) 1253–69. doi: 10.32604/cmc.2021.015655
10. Bhar A. *Is it Possible to Ensure COVID19 Vaccine Supply by Using Plants*?. Springer (2021).
11. Aslam B, Javed AR, Chakraborty C, Nebhen J, Raqib S, Rizwan M. Blockchain and ANFIS empowered IoMT application for privacy preserved contact tracing in COVID-19 pandemic. *Pers Ubiquitous Comput*. (2021) 1–17. doi: 10.1007/s00779-021-01596-3
12. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. *IEEE J Biomed Health Inform*. (2020) 24:2733–42. doi: 10.1109/JBHI.2020.3001216
13. Mondal A, Mahata S, Dey M, Das D. Classification of COVID19 tweets using machine learning approaches. In: *Proceedings of the Sixth Social Media Mining for Health (# SMM4H) Workshop and Shared Task*. Mexico City (2021). p. 135–7.
14. Ashokkumar P, Shankar GS, Gautham S, Reddy MPK, Reddy GT. A Two-Stage Text Feature Selection Algorithm for Improving Text Classification. *ACM Trans Asian Low-Resour Lang Inf Process*. New York, NY: Association for Computing Machinery (2021). p.19. doi: 10.1145/3425781
15. Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generat Comput Syst*. (2021) 117:47–58. doi: 10.1016/j.future.2020.11.022
16. Lopez CE, Vasu M, Gallemore C. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. *arXiv preprint* arXiv:200310359. (2020).
17. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*. (2021) 16:e0245909. doi: 10.1371/journal.pone.0245909
18. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res*. (2020) 22:e19016. doi: 10.2196/19016
19. Barkur G, Vibha GBK. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. *Asian J Psychiatr*. (2020) 51:102089. doi: 10.1016/j.ajp.2020.102089
20. Pedrosa AL, Bitencourt L, Fróes ACF, Cazumbá MLB, Campos RGB, de Brito SBCS, et al. Emotional, behavioral, and psychological impact of the COVID-19 pandemic. *Front Psychol*. (2020) 11:566212. doi: 10.3389/fpsyg.2020.566212
21. Fernandes N. *Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy*. Available at SSRN 3557504 (2020).
22. Abbasi A, Javed AR, Chakraborty C, Nebhen J, Zehra W, Jalil Z. ElStream: an ensemble learning approach for concept drift detection in dynamic social big data stream learning. *IEEE Access*. (2021) 9:66408–19. doi: 10.1109/ACCESS.2021.3076264
23. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. *Sci Rep*. (2020) 10:1–10. doi: 10.1038/s41598-020-73510-5
24. Samuel J, Ali G, Rahman M, Esawi E, Samuel Y, et al. Covid-19 public sentiment insights and machine learning for tweets classification. *Information*. (2020) 11:314. doi: 10.3390/info11060314
25. Mittal A, Patidar S. Sentiment analysis on twitter data: a survey. In: *Proceedings of the 2019 7th International Conference on Computer and Communications Management*. Bangkok (2019). p. 91–5.
26. Kumar DA, Chinnalagu A. Sentiment and emotion in social media covid-19 conversations: SAB-LSTM approach. In: *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*. Moradabad: IEEE (2020). p. 463–7.
27. Sciandra A. COVID-19 outbreak through tweeters' words: monitoring italian social media communication about COVID-19 with text mining and word embeddings. In: *2020 IEEE Symposium on Computers and Communications (ISCC)*. Rennes: IEEE (2020). p. 1–6.
28. Zhang S, Zheng D, Hu X, Yang M. Bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 29th Pacific Asia conference on language, information and computation*. Shanghai (2015). p. 73–8.
29. Mukherjee S, Malu A, Ar B, Bhattacharyya P. TwiSent: a multistage system for analyzing sentiment in twitter. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. New York, NY (2012). p. 2531–4.
30. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: a survey. *Wiley Interdiscipl Rev*. (2018) 8:e1253. doi: 10.1002/widm.1253
31. Prabha MI, Srikanth GU. Survey of sentiment analysis using deep learning techniques. In: *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. Chennai: IEEE (2019). p. 1–9.
32. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *IEEE Access*. (2020) 8:181074–90. doi: 10.1109/ACCESS.2020.3027350
33. Sanders A, White R, Severson L, Ma R, McQueen R, Paulo HCA, et al. Unmasking the conversation on masks: natural language processing for topical sentiment analysis of COVID-19 Twitter discourse. *medRxiv*. (2020) doi: 10.1101/2020.08.28.20183863

34. Nairn J, Ostendorf B, Bi P. Performance of excess heat factor severity as a global heatwave health impact index. *Int J Environ Res Public Health*. (2018) 15:2494. doi: 10.3390/ijerph15112494
35. Bandi A, Fellah A. Socio-Analyzer: a sentiment analysis using social media data. In: *Proceedings of 28th International Conference, Vol. 64.* (2019). p. 61–7.
36. Tang D, Wei F, Qin B, Yang N, Liu T, Zhou M. Sentiment embeddings with applications to sentiment analysis. *IEEE Trans Knowl Data Eng*. (2015) 28:496–509. doi: 10.1109/TKDE.2015.2489653

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with several of the authors AJ and ZJ.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Jalil, Abbasi, Javed, Badruddin Khan, Abul Hasanat, Malik and Saudagar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fpubh.2021.831404

A Recommendation System Based on AI for Storing Block Data in the Electronic Health Repository

Vinodhini Mani[1]*, *C. Kavitha*[1], *Shahab S. Band*[2]*, *Amir Mosavi*[3,4,5]*, *Paul Hollins*[6] and *Selvashankar Palanisamy*[7]

[1] Department of Computer Science and Engineering, School of Computing, Sathyabama Institute of Science and Technology, Chennai, India, [2] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan, [3] Faculty of Civil Engineering, TU-Dresden, Dresden, Germany, [4] Institute of Information Society, University of Public Service, Budapest, Hungary, [5] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [6] Cultural Research Development School of Arts, Institute of Management, University of Bolton, Bolton, United Kingdom, [7] Intelligent Automation, Ford Motors Pvt. Ltd., Chennai, India

The proliferation of wearable sensors that record physiological signals has resulted in an exponential growth of data on digital health. To select the appropriate repository for the increasing amount of collected data, intelligent procedures are becoming increasingly necessary. However, allocating storage space is a nuanced process. Generally, patients have some input in choosing which repository to use, although they are not always responsible for this decision. Patients are likely to have idiosyncratic storage preferences based on their unique circumstances. The purpose of the current study is to develop a new predictive model of health data storage to meet the needs of patients while ensuring rapid storage decisions, even when data is streaming from wearable devices. To create the machine learning classifier, we used a training set synthesized from small samples of experts who exhibited correlations between health data and storage features. The results confirm the validity of the machine learning methodology.

Keywords: artificial intelligence, machine learning, health repository, patients, health data, storage, deep learning

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Abdul Rehman Javed, Air University, Pakistan

***Correspondence:**
Vinodhini Mani
vinodhini.cse@sathyabama.ac.in
Shahab S. Band
shamshirbands@yuntech.edu.tw
Amir Mosavi
amir.mosavi@mailbox.tu-dresden.de

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *08 December 2021*
Accepted: *20 December 2021*
Published: *21 January 2022*

Citation:
Mani V, Kavitha C, Band SS, Mosavi A, Hollins P and Palanisamy S (2022) A Recommendation System Based on AI for Storing Block Data in the Electronic Health Repository. Front. Public Health 9:831404. doi: 10.3389/fpubh.2021.831404

INTRODUCTION

In the modern era, clinicians no longer manage health data exclusively, but are increasingly responsible for obtaining consent from patients (1). The rights of patient's access to, analysis of, and exchange of their health information have evolved dramatically (2). The majority of patients are dissatisfied with their health care providers after sharing self-tracking data (3). It is still possible to enhance patient health care by incorporating patient health data into the current health data systems. Literature has identified various categories of patient health information (4). These categories include information about medications, biometrics, behavioral information, data about social interactions, genetics, psychological data, data about symptoms, and reports. Blockchain-based interplanetary file system secondary storage of health data has been implemented to safeguard the privacy and security of patient health information (5). Yet very few studies have evaluated how patients' health data is stored. A key component of the proper management of health data is protecting the privacy and confidentiality of the patient while maintaining data accessibility for relevant stakeholders. Studies indicate that health data security poses a massive threat. This is evidenced by the proliferation of medical devices with limited memory and power (6, 7) and substantial medical data repositories (8). Many types of organizations are responsible for managing the massive amount of health data.

Health data is often portrayed as being sensitive to all patients with the same level of privacy and confidentiality; however, this is not true in practice because it is not equally sensitive to everyone at the same time. When a patient reaches a high level of public prominence, she may surrender the ECG data she generated on her own and to her cardiologist. This data can be accessed by other healthcare providers through an electronic health record. A patient who wishes to keep her pregnancy test results private may be forced to allow her provider to store her pregnancy test results. The dissemination of health data between multiple providers who manage data repositories now enables the storage medium to be customized based on patient needs. This includes the cost, size, security, confidentiality, and privacy of each chunk of data. Hybrid execution models, such as those described by the author (9), allow sensitive data to be stored in private clouds while no sensitive data is maintained in public clouds. Nevertheless, it does not specifically address health data processing. Communication between the two cloud platforms also takes time, and computations that rely on bandwidth use a lot of resources. A hybrid cloud platform was developed by (10) for solving this problem. Medical sensors, apps, and devices provide data to artificial intelligence, which enables the automatic diagnosis of health conditions. Health data, including ECG, blood pressure, and pulse rate, can be classified as normal or abnormal by algorithms based on a range of conditions and thresholds set by healthcare professionals. Clinical research and clinical care are usually aided by abnormal data. Using the Body Area Sensor Network, (8) developed an agent-based system developed for elderly people to preserve abnormal data. Health information is generated in enormous quantities nowadays, so a diverse storage solution is needed (11). Several researchers have examined the performance and cost parameters of various Cloud Service Providers (CSPs) to design methods for selecting suitable CSPs for storing consumers' data (12–14). High-performance cloud services minimize the time spent in operations but incur high costs. Additionally, researchers are investigating blockchain technology for its promise of security and privacy for health data management. Combining blockchain-based eHealth with traditional health databases is possible, which can be arranged based on users' preferences and the possibility of utilizing the data in the future. However, due to the design of blockchains, they are not suitable for hosting large amounts of health data. A software agent that knows the patient's preferences is inserted inside the application in (15). Nonetheless, they never described a way to make this decision. To assist in choosing storage repositories, we developed a model that incorporated not only (8)'s criteria, but also aspects like data confidentiality, privacy, and quality of performance.

Motivation

Every Blockchain miner owns a local ledger, so this technology allows transactions to be verified and processed without the need for third parties. Verifying transactions does not require a centralized server. Document alterations cannot be guaranteed through conventional database storage and blockchain-based hash management. Data is only detectable in a blockchain if a hash pointer holds a pointer to it. Depending on the patient, personal preferences, and other factors, the sensitivity and significance of the health information are also different from repository to repository. Choosing the right repository is extremely crucial. As wearable sensors continuously stream health data, the challenges are exacerbated. In (16), the author has surveyed the importance of artificial intelligence in healthcare. The prediction of COVID-19 infected patients using artificial intelligence has been implemented in (17), but there is a need for an appropriate repository to store the data.

Contribution

In our research, we considered the variation in data sensitivity, volume, and other factors to locate the appropriate system to manage health records. The flow diagram of the paper contribution is shown in **Figure 1**. Collect the health data and health repository parameters. Evaluations of both health information and health repository parameters are given a score. The machine learning-based recommendation model for health data storage proposes a way to distribute health data among multiple repositories. A model for automated health data storage recommendation is being developed to determine appropriate storage repositories. Through correlation analysis, user preferences, and clinical heuristics, a machine learning-based classifier is used to map health data characteristics to each repository. Patients' security and privacy preferences are taken into account as well as the sensitivity of health data.

Organization

Following are the sections of the paper: section Background addresses related work. In section Model for Recommendation of Health Repositories, we present the proposal for a recommendation model for a health repository. Section Implementation describes how the system will be implemented. The results and evaluation of performance will be discussed in section Results and Discussion. Conclusions and future work will be discussed in section Conclusion.

BACKGROUND

Big Data cannot be stored, accessed, or analyzed with a single health record system. Patients can lose medical information when their electronic health records are malfunctioning (18). Due to the manual uploading of data generated by wearable sensors to personal health records, caregiver responses were delayed. For this reason, (19) developed methods for storing patient-generated health information on commercial blood glucose monitors. The electronic health record system could be made to fit the streamed data if it is filtered or compressed (20). In (21–24), a number of action plans and standards were advocated for the adoption of an electronic health record system. A selection of an electronic health record should take into account functional requirements, troubleshooting, and optimization features (22). The author provides a list of steps to follow before buying an electronic health record system. Checklists mostly cover client meetings on site, site visits, and maintaining live workflows. Health data sources such as hospitals, clinics, insurers, and patients should be integrated into

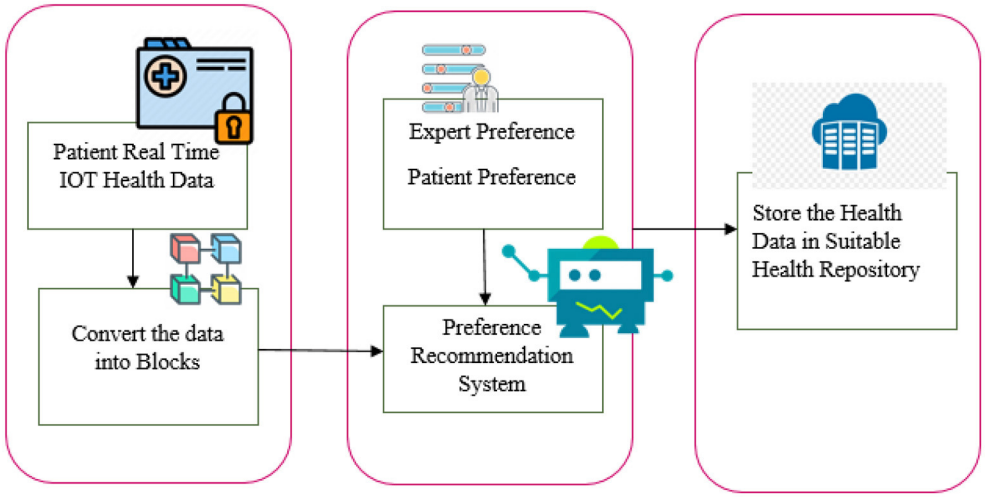


FIGURE 1 | Paper contribution flow diagram.

centralized databases, according to the author (25). In particular, patient-centered health data with high degrees of structural heterogeneity must be stored and processed quickly because of their high volume and rate. For health data, to provide useful insights, precision is essential, but some sources produce vague and inaccurate information. Distributed data storage systems do offer some relief to these issues (26). Various cloud storage mediums have been examined. A machine learning and deep learning model is used to predict the thermal sensation vote system (27). Utilization of a compression algorithm to retrieve the health repository data as fast as possible using blockchain and interplanetary file systems (IPFS) without data loss (28). Diabetic Retinopathy is efficiently classified using a deep learning and machine learning algorithm (29). Genetic algorithm with fuzzy logic is a tool to help medical practitioners diagnose heart disease at an early stage using adaptive genetic algorithm with fuzzy logic (AGAFL) (30). Health data storage systems and data properties were not considered in the selection of repositories. Furthermore, no machine learning mechanisms were developed to cater to user preferences.

In the next section, we describe how we facilitate distributed health data management.

MODEL FOR RECOMMENDATION OF HEALTH REPOSITORIES

As data streams increase, the need for storage decisions becomes more frequent, making manual consultation with patients an inefficient process that requires an automated solution. It is, however, impossible to prespecify the data storage requirements for each patient that will apply to all possible future contexts. The learning classifier may generalize to a broader range of mappings based on a manual mapping specification by an expert.

The following sections explain in detail the overall approach described in **Figures 2**, **3**. Data storage requirements - an illustration of which is displayed in layer 1 of **Figure 2**, consists of a set of variables or features that characterize the requirements for storing a chunk of data. Some of the attributes' values have been shown to be numerical [1–10] and others to be qualitative. Secondly, each instance of the dataset contains the specifications required to store each chunk of data as shown in **Figure 2**.

Health Repository Evaluation Criteria are calculated in layer 3 by adding a rating provided by an expert group. These criteria reflect the characteristics of storage repositories as shown in **Figure 2**. Three standards apply to rank five storage repositories. Medical professionals and patients themselves may create clinical heuristic rules in layer-3 of **Figure 2** and each instance in the dataset is categorized according to the preferences of the users. A storage repository can be assigned to an instance based on heuristic rules in a real-world situation. The correlation coefficient offers an inference of a class label when preferences and heuristics do not match well. The health repository requirements can be mapped to layer-4 (user and expert expectations) by a machine learning classifier, as shown in **Figure 2**. In **Figure 3**, a recommendation framework for health repositories is illustrated. There are two parts to the framework: determining which standards should be used for the storage and assessment of data and implementing machine learning.

IMPLEMENTATION

This recommendation system assumes that a patient is in full control of his or her decision regarding storage. It is impossible to make decisions manually in many cases because they are made so frequently. Hence, automated processes are essential. In the mapping process, the characteristics of a repository managed by an agent group are matched with the characteristics of data about the storage requirements of patients. Because patients' storage requirements vary so much, it is impossible to predetermine every possible scenario. By utilizing a set

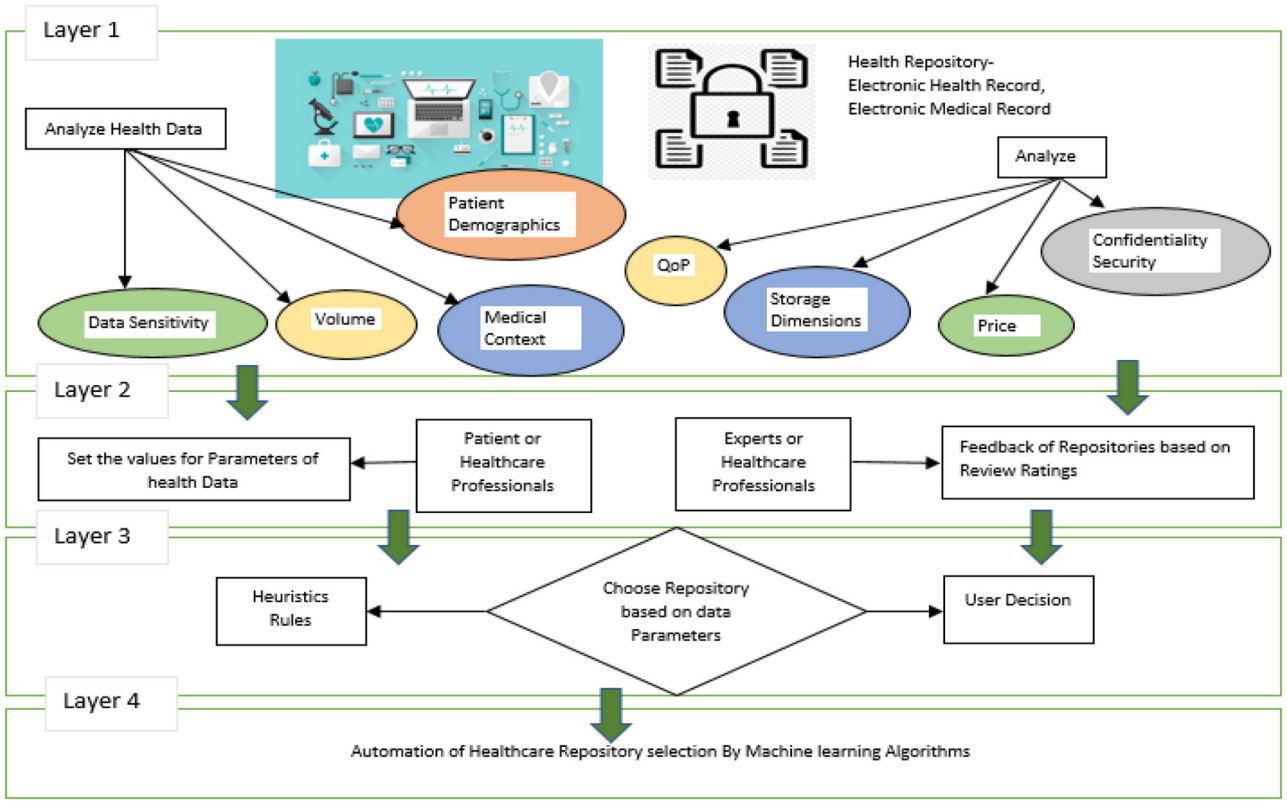


FIGURE 2 | Proposed system architecture.

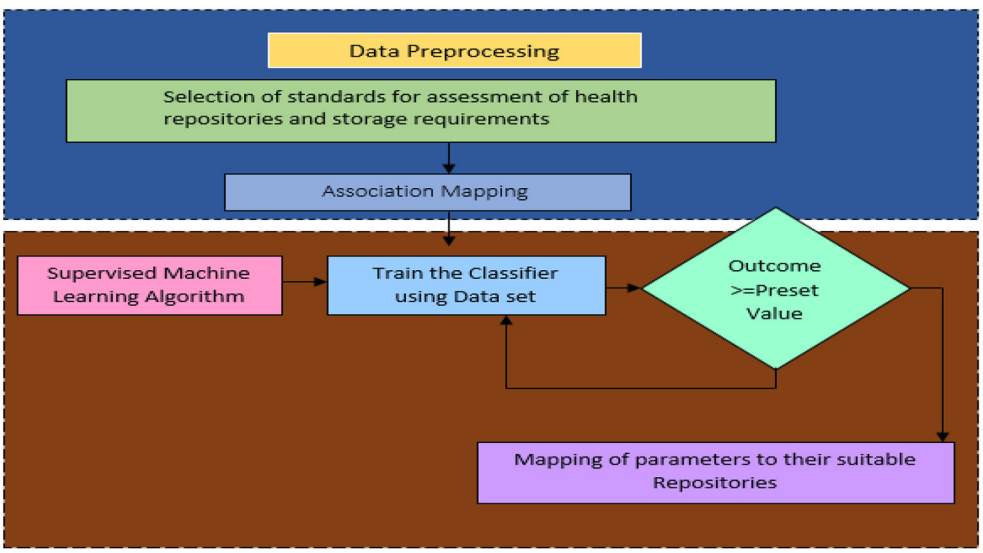


FIGURE 3 | Proposed health repository recommendation system.

of mappings that is specified manually by experts, machine learning is used to generalize a mapping over a wide range of patient contexts. This methodology involves defining a set of attributes that describe what chunk of data needs to be stored. There are numerical values and categorical values assigned to those attributes. Thus, a dataset containing these attributes will be created, with each instance representing a different set of storage requirements. A group of experts' ratings are then

used to determine the characteristics of the available storage mediums. To determine what class each instance falls into, statistical correlation and heuristic rules are employed. Based on the training datasets, the supervised machine learning classifier maps the data into a storage repository. **Figure 3** illustrates two components of the recommendation system: Data Pre-processing and Supervised Machine Learning. According to **Figure 3**, the upper portion of the framework contains the characteristics of the data storage requirements. There are a number of features that demonstrate the characteristics of health repositories. A number of associations were found between the two groups of features.

Data Preprocessing

The data collected from hospitals and patients undergoes a preprocessing process, which includes analyzing data storage requirements, identifying sensitive data areas, analyzing the volume of each record, analyzing the patient health profile, determining the demographics of patients, and analyzing health repository parameters as well as storage, cost, security, privacy, and performance.

Characteristics of Data Storage Requirements

To determine which repository is the best option, consideration is given to the sensitivity of the data, the volume of the data, medical care context, and demographics of the patient.

Sensitivity of the Data

It is imperative to prevent unauthorized access to all health-related data. Depending on the data type, some breaches are more likely than others. Depending on the individual's preferences and context, the level of data sensitivity may vary.

The Volume of the Data

Reports, medical diagnoses, and medication summaries are not frequently created, which means that their storage needs are less than those of health data sets.

Context of Medical Care

The context may be palliative care, critical care, chronic illness, or no chronic illness. The context may also differ based on the country.

Demographics of Patients

Several factors can play a significant role in determining which storage medium to use, such as socioeconomic status, occupation, education, and nationality.

Health Repository Evaluation Parameters

Evaluation parameters for health repository such as security, privacy, cost, storage capacity, and performance. **Table 1** shows the parameters and criteria of the health repository evaluation.

The Relationship Between Repository Evaluation Standards and Data Features

Medical records, in particular those generated by patients, are to be transferred to a health record system that reflects the preferences of the user and the data requirements. Health data requirements and criteria for evaluating storage are correlated in a one-to–to-many fashion as implemented in **Algorithm 1**. Some associations are strong, and some are weak. To facilitate the rapid processing of highly confidential data, a health record system may accept data blocks in plaintext format. Data with relatively low confidentiality can be highly sensitive due to the demographic characteristics of patients. Data about a patient's demographics, such as their educational background and professional experience, may affect their privacy concerns. Users can then choose from a variety of storage repositories that

TABLE 1 | Health repository evaluation.

Assessment parameters	Survey questions for health repository ratings
Storage	Can the repository be used to store Big Data?
	Regarding processing Big Data, what is the repository's role?
	Are there any benefits to storing continuously streamed data in the repository?
Cost	Does deployment cost a lot?
	Does maintenance cost much?
	What is the service cost?
Security	Is the storage repository capable of maintaining data integrity?
	Does the storage repository have 24/7 accessibility?
	Are storage repositories resistant to cyberattacks?
Privacy	Is data accessible to third parties?
	Is the access control right given to the owner of the health records?
Performance	How fast can you upload files?
	Is it possible to retrieve data quickly?
	Is it possible to process data quickly?

Algorithm 1: Association mapping ().

```
Step 1:    Begin
Step 2:    Let Data Source as DS;
Step 3:    Let Storage Requirements as SR;
Step 4:    Let Health Repository Parameters as HRP;
Step 5:    For each data ϵ DS do
Step 6:       For each Storage Requirement ϵ SR do
Step 7:          Collect the data;
Step 8:          Identify the SR;
Step 9:          Collect the HRP;
Step 10:      For each SR and HRP do
Step 11:         Analyze the parameters using Evaluation
                 Criteria;
Step 12:         If (SR ϵ HRP)
Step 13:               SR (SR1…n)→HRP (HRP1…n);
Step 14:               Create Association Dataset as AD;
Step 15:         Else
Step 16:               Print Not Associated;
Step 17:         End; End; End; End; End;
```

Table 2 | Association mapping.

S. No	Characteristics of data storage requirements	S. No	Health repository evaluation parameters	Association mapping
1	Sensitivity of the data	A	Storage	1→(B,C,D,E)
2	The volume of the data	B	Cost	2→(A)
3	Context of Medical Care	C	Security	3→(E)
4	Demographics of patients	D	Privacy	4→(B,C,D,E)
		E	Performance	

Algorithm 2: Health repository recommendation system ().

```
Step 1:   Begin
Step 2:   data collected from various data sources;
Step 3:   Call Association Mapping ();
Step 4:   For each Health Data Block ϵ HB do
Step 5:     Select the Supervised Machine learning algorithm;
Step 6:     Train the Data block HB;
Step 7:     Apply Heuristic Rule;
Step 8:     If (Accuracy ≥ Threshold)
Step 9:       Test data;
Step 10:      Allocate the Health Data Block
              HB→Health Repository HR;
Step 11:      Send (Recommend Repository to Patients);
Step 12:      Break;
Step 13:    Else
Step 14:      Continue;
Step 15:    End; End; End;
```

protect their confidentiality. The sample association mapping as shown in **Table 2**.

Supervised Machine Learning Algorithm

Dynamically suggest health repositories based on supervised learning for particular data blocks, which is implemented using **Algorithm 2**. A training dataset must be generated for every instance of the dataset in addition to the labeled training datasets. Health repositories will be assigned data blocks that have a number of attributes. Among the attributes are some that are directly linked to the data block and others that are directly linked to the patient. Attributes include data sensitivity, volume, context of care, and demographics of the patients. The health repository should consider for evaluation such as electronic health records, cloud based electronic health records, blockchain based electronic health records, patient health record, and Electronic Medical Records. We considered the following health repository parameters in this study: security, privacy, cost, storage capacity, and performance. Each repository has been assigned a rating value ranging from 1 to 10.Whenever other attributes are not significant in determining the health repository, a linear regression Y (1) is calculated to label the instance as shown in Equation 1.

$$Y = A + RX \tag{1}$$

$$\mathbf{R} = \mathbf{n}\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{xiyi}\right) - \left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{xi}\right)\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{yi}\right) \tag{2}$$

$$\mathbf{A} = \frac{\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{yi}\right) - \mathbf{R}\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{xi}\right)}{\mathbf{n}} \tag{3}$$

Where R is the Coefficient which contains R1,R2,R3,R4,R5,R6,R7,R8,R9,R10 are calculated between the set of data storage requirements(DR) as shown in equation 2. Here are the evaluation criteria for Electronic health record (D1), Patient health record (D2), Cloud-based electronic health record (D3), Blockchain-based electronic health records (D4), and Electronic Medical records (D5). The calculation of health repository recommendation Di is estimated using the equation:

$$\boldsymbol{Di = High\ (R1,\ R2\ldots..Rm)} \tag{4}$$

M is the number of health repositories and n is the rating criteria. Secondly, the choice of a health data repository can be influenced by the decision of the healthcare professional, the preferences of the user, and a variety of factors such as normal or abnormal behavior patterns and patient health status, as well as other demographic factors. Patients with unusual health patterns should store their health records in a repository that health care professionals can access quickly. A less secured and less expensive repository can be used to store data which is hardly ever accessed by health care professionals. Different users may have different privacy preferences, and those preferences may change over time based on different contexts (31). The health record system for a patient should take into account a variety of factors. There are several factors involved, such as medical conditions, personal characteristics, socioeconomic status, as well as the type and significance of data. The level of privacy and security preferences of individuals may change over time as well. In contrast to patients with terminal illnesses, young individuals may be more concerned with privacy and security. By considering author preference, some of the sample user preference and health professional preference heuristic rules were implemented, as shown below:

1. If (Data= standard && volume=large)

   Then
   Storage Repository=Cloud based Health Record Management System

2. If (Data= standard && volume=low)

   Then
   Storage Repository=Blockchain enabled Personal Health Record System

3. If (Data=Unusual patterns && volume=low)

Table 3 | Mapped sample training data set.

Information block	Sensitivity data	Volume	Context of medical care	Social status	Profile visibility	Patient status	Health repository
Data Block 1	1	2	3	3	high	Typical	Blockchain based electronic health record
Data Block 2	2	5	3	5	Low	Typical	Cloud electronic health record
…….	…	…	…	….	…	….	…..
Data Block n	3	2	3	2	1	Abnormal	Electronic medical record

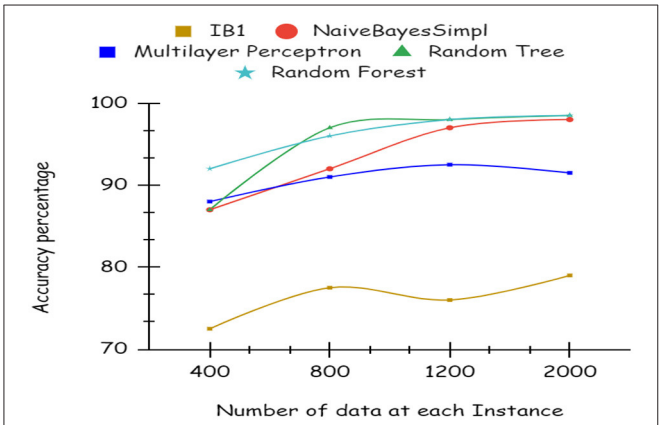


FIGURE 4 | Accuracy using 10-fold cross validation.

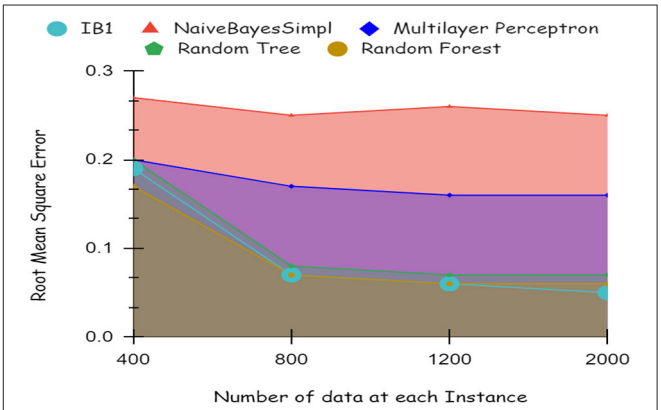


FIGURE 5 | RMSE using 10-fold cross validation.

Then
Storage Repository=Blockchain based Electronic Medical Record

4. If (Patient= Famous Personality && health condition = Good)

Then
Storage Repository=Blockchain based Electronic Health Record

5. If (Patient= Famous Personality && health condition = Serious)

Then
Storage Repository=Blockchain based Electronic Medical Record

6. If (Data of type Disease)

Then
Store data in Disease Registry

RESULTS AND DISCUSSION

Research was conducted on supervised machine learning classification techniques. Using the WEKA tool, different classification algorithms were tested. The study used an Intel Core i7 6700H processor with up to 3.5 GHz and 16 GB of RAM. The dataset was divided into training and test sets. Data preprocessing is performed prior to analysis. To train the data in the recommended health repository, linear regression data blocks and user and health professional preference rules have been used. During this experiment, we determine whether the classifiers can learn how to classify data distributions. The training datasets each

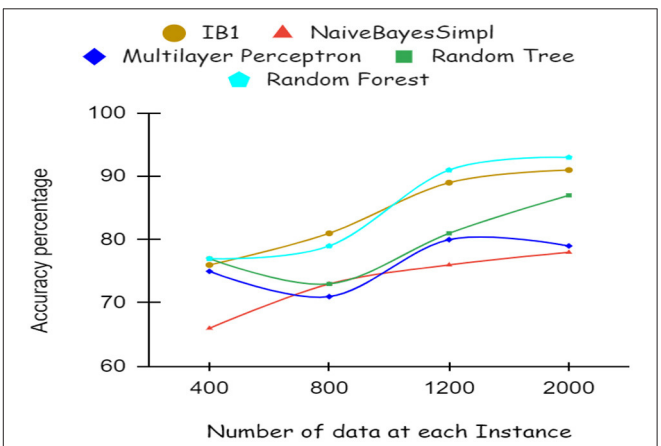


FIGURE 6 | Accuracy of percentage split dataset.

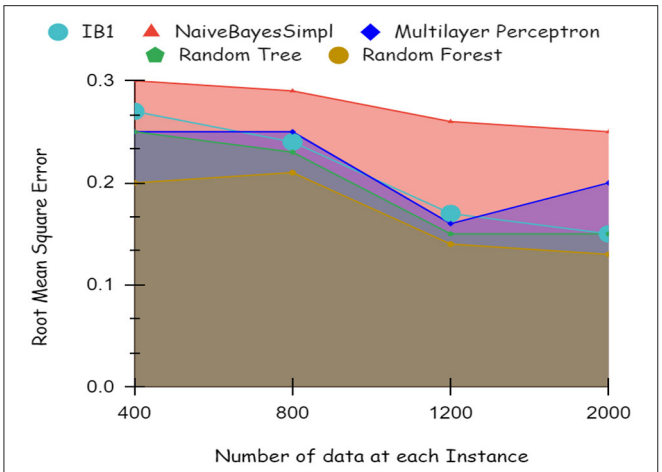


FIGURE 7 | RMSE of percentage split dataset.

FIGURE 8 | Performance loss of training and test set.

contain 400, 800, 1200, and 2000 instances. **Table 3** shows the mapped sample training dataset.

Four different classifiers were run on four datasets to test whether a machine learning algorithm could choose an appropriate storage medium, NaïveBayesSimple, Multilayer Perceptron, Random Forest Classifier, Random Tree and the IB1 algorithm are four different types of classifiers trained here. Several classification techniques were compared using Python to determine their accuracy scores (32).

Classification Model Accuracy

1. Confusion matrix.
2. Classification measure.

Confusion Matrix

In the confusion matrix, N is the number of target classes, and N is the number of rows. It is used to evaluate the performance of a classification model. Machine learning is used to predict target values from the actual values in the matrix. True Positive (TP) and True Negative (TN) rates should be high and False Positive (FP) and False Negative (FN) rates are low for a successful model. A confusion matrix as is always more appropriate as a machine learning model evaluation criterion when working with an imbalanced dataset.

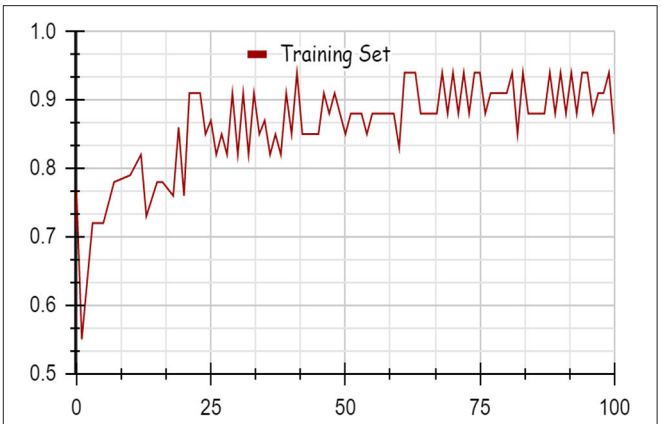


FIGURE 9 | Performance accuracy of training and test set.

Classification Measure

As an evaluation measure, the classification measure is used in addition to the confusion matrix. They are:

1. Accuracy.

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}} \quad \textbf{0.0} < \textbf{Accuracy} < \textbf{1.0} \tag{5}$$

2. Precision.

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP+FP}} \tag{6}$$

3. Recall.

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP+FN}} \tag{7}$$

4. F1-Score.

$$\textbf{F1-Scrore} = 2\frac{\textbf{Recall}^{*}\textbf{Precision}}{\textbf{Recall+ Precision}} \tag{8}$$

5. Sensitivity and specificity.

$$\textbf{Sensitivity} = \frac{\textbf{TP}}{\textbf{TP+FN}} \tag{9}$$

$$\textbf{Specificity} = \frac{\textbf{TN}}{\textbf{TN+FP}} \tag{10}$$

6. Root mean square error.

Modified Mean Square Error (MSE) is a variation of Root Mean Square Error (RMSE). Measuring the mean square error squared is equivalent to this metric. The RMSE of an ideal model is zero, just as the MSE and MAE are zero.

$$RMSE = \sqrt{\frac{1}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\textbf{Actual Values} - \textbf{Predicted Values})^{2}} \tag{11}$$

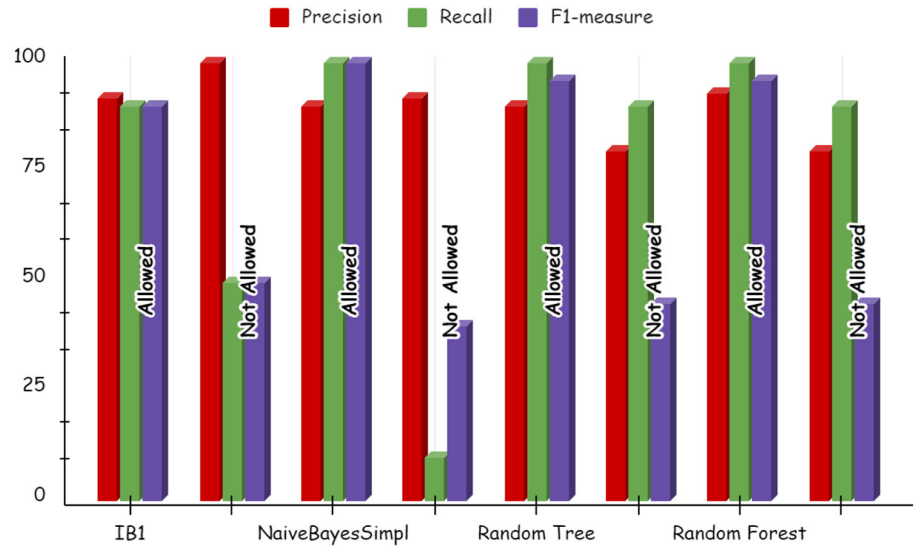


FIGURE 10 | Deep learning results for cloud electronic health record.

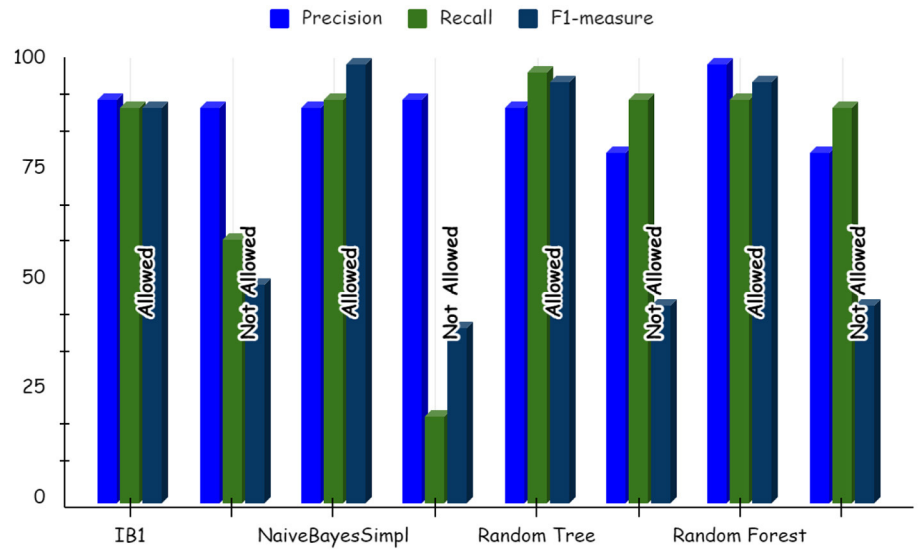


FIGURE 11 | Deep learning results for blockchain based electronic health record.

Result Analysis

As illustrated by the graph in **Figure 4**, Random Forest classifiers become more accurate as the number of instances increases, as shown by a 10-fold cross-validation analysis. A balanced ratio of each class was found in the dataset of 1,200 records, thus all classifiers performed better. The Random Forest performed best, with 98.21% accuracy. On the 2,000-record dataset, however, all classifiers had lower accuracy, largely because the dataset was skewed. Compared to other classifiers, Random Forest exhibits lower root mean square error in **Figure 5**. **Figure 6** illustrates the percentage split results, which are less accurate than the cross-validation results presented in 10-fold cross-validation.

By using a percentage split, 80% of the data were used for training and 20% for testing. The classifier is trained only once, as seen in **Figure 7**, which demonstrates low accuracy and large RMSE. Artificial intelligence is a technique for deep learning.

Using deep learning networks, unstructured or unlabeled data can be learned unsupervised. Real-world health repositories are usually recommended based on unstructured and unlabeled datasets. For our synthetic dataset, we analyzed the accuracy using a deep learning algorithm. A deep learning model is run on the synthetic dataset, and it shows 88.70 percent accuracy. It is implemented in Python. There are three hidden layers in the model; the first of these layers has 100 output nodes,

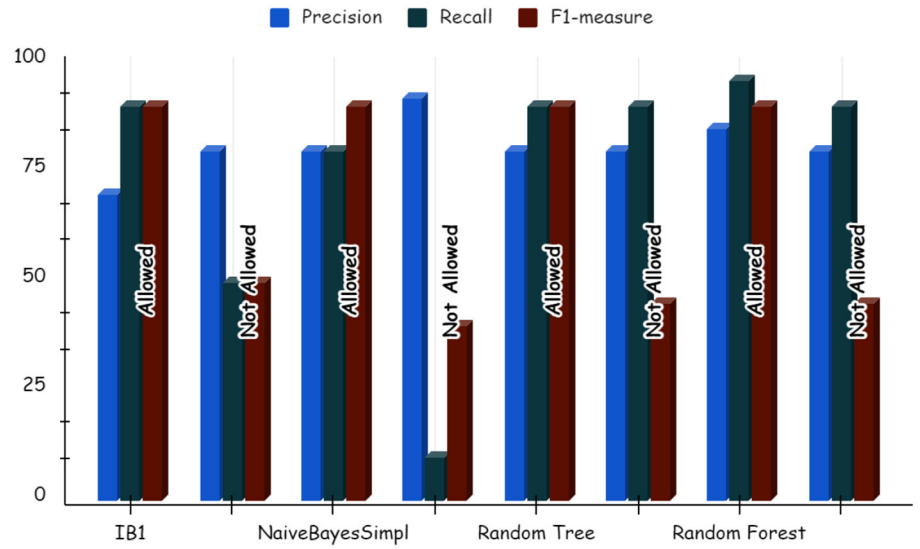


FIGURE 12 | Deep learning results for electronic medical record.

while the second and third have five output nodes each. Training is done with 100 iterations and eight batches are used. The training dataset is shown in **Figures 8**, **9**, with a Y-axis showing the loss and X-axis showing the number of iterations. A deep learning classifier and a machine learning classifier are displayed in **Figures 10–12** for the classification. With reference to recall, F1-measure, and precision, the Random Forest classifier outperformed the other tested classifiers. Classes that were allowed and those that were not were included in the experiment. In terms of recall, precision, and F1-measure, the Random classifier scored 93, 100, and 96% for cloud electronic health records, 100, 92, and 96 for blockchain-based electronic health records, and 85, 96, and 90 for electronic medical records. In terms of the allowed class, the rest of the experimented models perform well. In terms of the disallowed class, they did not perform well.

The accuracy of the classifier supports the use of machine learning to map the health storage mediums to health data blocks. Given the growing volume of health data that will need to be stored and accessed globally, this machine learning model may play a crucial role in improving storage and access arrangements in the future. This will make health data storage easy and straightforward for consumers. In addition, they would be able to ensure that the size of the data store is manageable. It can help to determine which storage solution best fits the requirements of different data assets using a machine learning model.

Mapping of Health Data Parameters to Repositories

Medical technology is expected to develop health record systems in the future. Health records are taking on novel forms as a result of the expansion of medical data. As described below, the proposed system will support various data variations and health records. First, the system requests the ratings for the latest health record on the basis of health parameters from the IT staff and healthcare professionals. Second, the system relabels instances from the entire training dataset. As soon as a new instance is created, the old instances' labels do not change.

CONCLUSION

Health data will increasingly be preserved in a variety of repositories, so patients can select the repository that best meets their needs. Patients are realistically expected to avoid using a single repository for all their health data because the context of treatment, patterns of data, and legal constraints may change. To automate the storage decision, a selection algorithm must be developed. This is especially relevant in the case of constantly streaming health data. The process of choosing the right repository is complicated. In addition to knowledge of storage features used for interoperability, data security, and privacy, regulatory concerns must also be considered. To preserve confidentiality, we propose distributing health data among various vendors. By keeping medical records together, confidentiality will also be preserved. Based on factors like data type, sensitivity level, significance, patient safety, and privacy requirements, this model can recommend which health data blocks should be stored on which storage medium. When applied to the dataset generated, random forest yielded the highest accuracy of 96.4%. Accuracy of algorithms depends on the dimension, origin, and nature of the data. As a result, we intend to evaluate these various algorithms with different characteristic datasets in the near future. In the future, we will implement a role-based access control system to store medical record information by

integrating the health repository recommendation system to allow access to the health records based on the permission of patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

VM and CK: conceptualization, methodology, investigation, data curation, and writing—original draft preparation. SB, AM, and PH: software, validation and visualization, and resources. VM, AM, and PH: formal analysis. VM, SB, and AM: writing—review and editing and supervision. VM, CK, SB, AM, and PH: project administration. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Plastiras P, O'Sullivan D. Exchanging personal health data with electronic health records: A standardized information model for patient generated health data and observations of daily living. *Int J Med Inform.* (2018) 120:116–25. doi: 10.1016/j.ijmedinf.2018.10.006
2. Cortez A, Hsii P, Mitchell E, Riehl V, Smith P. *Conceptualizing a data infrastructure for the capture, use, and sharing of patient-generated health data in care delivery and research through 2024 (white paper).* (2018).
3. Chung CF, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient generated data. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, (2016). p. 770–786. doi: 10.1145/2818048.2819926
4. Lordon RJ, Mikles SP, Kneale L, Evans HL, Munson SA, Backonja U, et al. How patient-generated health data and patient-reported outcomes affect patient–clinician relationships: A systematic review. *Health Inform J.* (2020) 26:2689–706. doi: 10.1177/1460458220928184
5. Mani V, Manickam P, Alotaibi Y, Alghamdi S. Hyperledger Healthchain: patient-centric IPFS-based storage of health records. *Electronics.* (2021) 10:3003. doi: 10.3390/electronics10233003
6. Albahri A, Zaidan A, Albahri O, Zaidan B, Alsalem M. Real-time fault-tolerant mhealth system: Comprehensive review of healthcare services, opens issues, challenges and methodological aspects. *J Med Syst.* (2018) 42:137. doi: 10.1007/s10916-018-0983-9
7. Isern D, Moreno A. A systematic literature review of agents applied in healthcare. *J Med Syst*. (2016) 40:43. doi: 10.1007/s10916-015-0376-2
8. Vaidehi V, Vardhini M, Yogeshwaran H, Inbasagar G, Bhargavi R, Hemalatha CS. Agent based health monitoring of elderly people in indoor environments using wireless sensor networks. *Procedia Comput Sci.* (2013) 19:64–71. doi: 10.1016/j.procs.2013.06.014
9. Ko SY, Jeon K, Morales R. The hybrex model for confidentiality and privacy in cloud computing. *HotCloud.* (2011) 11:8. doi: 10.5555/2170444.2170452
10. Zhang H, Ye L, Du X, Guizani M. Protecting private cloud located within public cloud. In *Global Communications Conference (GLOBECOM).* IEEE, (2013). p. 677–681.
11. Stranieri A, Balasubramanian V. Remote patient monitoring for healthcare: a big challenge for big data. In *Managerial Perspectives on Intelligent Big Data Analytics*. IGI Global, (2019). p. 163–179. doi: 10.4018/978-1-5225-7277-0.ch009
12. Ruiz-Alvarez A, Humphrey M. A model and decision procedure for data storage in cloud computing. In *Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012).* IEEE Computer Society, (2012). p. 572–579. doi: 10.1109/CCGrid.2012.100
13. Ruiz-Alvarez A, Humphrey M. Toward optimal resource provisioning for cloud mapreduce and hybrid cloud applications. In: *Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing*. IEEE Computer Society, (2014). p. 74–82. doi: 10.1109/BDC.2014.14
14. Yoon MS, Kamal AE. Optimal dataset allocation in distributed heterogeneous clouds. In: *2014 IEEE Globecom Workshops (GC Wkshps)* IEEE, (2014). p. 75–80. doi: 10.1109/GLOCOMW.2014.7063389
15. Zhang Q, Lu J. Artificial intelligence in recommender systems. *Complex Intell Syst.* (2021) 7:439–57. doi: 10.1007/s40747-020-00212-w
16. Yang Y, Chen T. Analysis and visualization implementation of medical big data resource sharing mechanism based on deep learning. *IEEE Access.* (2019) 7:156077–88. doi: 10.1109/ACCESS.2019.2949879
17. Stock C, Dias S, Dietrich T, Frahsa A, Keygnaert I. Editorial: How can We Co-Create Solutions in Health Promotion with Users and Stakeholders? Front. *Public Health.* (2021) 9:773907. doi: 10.3389/fpubh.2021.773907
18. Andy YY, Shen CP, Lin YS, Chen HJ, Chen AC, Cheng LC, et al. Continuous, personalized healthcare integrated platform. In *TENCON 2012 IEEE Region 10 Conference*. IEEE, (2012). p. 1–6. doi: 10.1109/TENCON.2012.6412226
19. Peleg M, Shahar Y, Quaglini S, Fux A, García-Sáez G, Goldstein A, et al. Mobiguide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestationaldiabetes domains. *User Model User-Adapt Interact*. (2017) 27:159–213. doi: 10.1007/s11257-017-9190-5
20. Hohemberger R, da Rosa CE, Pfeifer FR, da Rosa RM, de Souza PS, Lorenzon AF, et al. An approach to mitigate challenges to the electronic health records storage. *Measurement.* (2020) 154:107424. doi: 10.1016/j.measurement.2019.107424
21. Busis NA. How can i choose the best electronic health record system for my practice? *Neurology.* (2010) 75:S60–4. doi: 10.1212/WNL.0b013e3181fc9888
22. Weathers AL, Esper GJ. How to select and implement an electronic health record in a neurology practice. *Neurol Clin Pract.* (2013) 3:141–8. doi: 10.1212/CPJ.0b013e31828d9fb7
23. Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten simple rules for digital data storage. *PLoS Comput. Biol*. (2016) 12:10. doi: 10.1371/journal.pcbi.1005097
24. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. *PLoS Comput Biol.* (2017) 13:e1005510. doi: 10.1371/journal.pcbi.1005510
25. Khan SI, Hoque ASML. Towards development of health data warehouse: Bangladesh perspective. In *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).* IEEE (2015). p. 1–6. doi: 10.1109/ICEEICT.2015.7307514
26. Mackey TK, Kuo TT, Gummadi B, Clauson KA, Church G, Grishin D, et al. 'Fit-for-purpose?'–challenges and opportunities for applications of blockchain technology in the future of healthcare. *BMC Med*. (2019) 17:68. doi: 10.1186/s12916-019-1296-7
27. Rehman SU, Javed AR, Khan MU, Nazar Awan M, Farukh A, Hussien A. PersonalisedComfort: a personalised thermal comfort model to predict thermal sensation votes for smart building residents. *Enterpr Inf Syst*. (2020). 1852316. doi: 10.1080/17517575.2020.1852316
28. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR. Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm. *J Circ Syst Comput.* (2021) 2250010. doi: 10.1142/S0218126622500104
29. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava, G. (2020). Deep neural networks to predict diabetic retinopathy. *J Ambient Intell Human Comput.* (2020) 1–14. doi: 10.1007/s12652-020-01963-7

30. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell.* (2020). 13:185–96.
31. Trojer T, Katt B, Schabetsberger T, Mair R, Breu R. The process of policy authoring of patient-controlled privacy preferences. In: *International Conference on Electronic Healthcare*. Springer, (2011). p. 97–104. doi: 10.1007/978-3-642-29262-0_14
32. Analytics V. What is confusion matrix (2020). Available online at: https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5 (accessed November 17, 2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mani, Kavitha, Band, Mosavi, Hollins and Palanisamy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fpubh.2021.792124

An Empirical Model to Predict the Diabetic Positive Using Stacked Ensemble Approach

Sivashankari R.[1], *Sudha M.*[1], *Mohammad Kamrul Hasan*[2*], *Rashid A. Saeed*[3], *Suliman A. Alsuhibany*[4] *and Sayed Abdel-Khalek*[5,6]

[1] School of Information Technology and Engineering, Vellore Institute of Technology (VIT), Vellore, India, [2] Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia, [3] Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, [4] Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia, [5] Mathematics and Statistics Department, College of Science, Taif University, Taif, Saudi Arabia, [6] Mathematics Department, Sohag University, Sohag, Egypt

Today, disease detection automation is widespread in healthcare systems. The diabetic disease is a significant problem that has spread widely all over the world. It is a genetic disease that causes trouble for human life throughout the lifespan. Every year the number of people with diabetes rises by millions, and this affects children too. The disease identification involves manual checking so far, and automation is a current trend in the medical field. Existing methods use a single algorithm for the prediction of diabetes. For complex problems, a single model is not enough because it may not be suitable for the input data or the parameters used in the approach. To solve complex problems, multiple algorithms are used. These multiple algorithms follow a homogeneous model or heterogeneous model. The homogeneous model means the same algorithm, but the model has been used multiple times. In the heterogeneous model, different algorithms are used. This paper adopts a heterogeneous ensemble model called the stacked ensemble model to predict whether a person has diabetes positively or negatively. This stacked ensemble model is advantageous in the prediction. Compared to other existing models such as logistic regression Naïve Bayes (72), (74.4), and LDA (81%), the proposed stacked ensemble model has achieved 93.1% accuracy in predicting blood sugar disease.

Keywords: random forest, KNN classifier, decision tree, gradient boosting, SVM and Gaussian Naïve Bayes, PIMA dataset, healthcare systems

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Ebuka Ibeke,
Robert Gordon University, United Kingdom
Senthilkumar Mohan,
VIT University, India
Naeem Ahmad,
National Institute of Technology Raipur, India

***Correspondence:**
Mohammad Kamrul Hasan
mkhasan@ukm.edu.my;
hasankamrul@ieee.org

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 October 2021*
Accepted: *30 November 2021*
Published: *21 January 2022*

Citation:
R. S, M. S, Hasan MK, Saeed RA, Alsuhibany SA and Abdel-Khalek S (2022) An Empirical Model to Predict the Diabetic Positive Using Stacked Ensemble Approach. Front. Public Health 9:792124. doi: 10.3389/fpubh.2021.792124

INTRODUCTION

People's regular foods contain a vast amount of carbohydrates and calories. Three primary reasons that a person may suffer from diabetes are genetics, lifestyle, and environment. The first reason for diabetic positivity is genetics. Family studies proved that the children whose parents are type 2 [Muoio and Newgard (1)] diabetic have three times more chances to develop diabetic positive than the parents who don't have diabetic positive. Lifestyle is the second reason for the diabetic positive because proven studies show that the individual lifestyle causes diabetic positive even though their ancestors are not diabetic positive. The third reason for the diabetic positive is adopting intricate weight loss mechanisms. It causes kidney failure or heart issues that lead to diabetes positive in the future.

The symptoms for the diabetic positive are eye powerlessness, sudden weight loss, frequent urination, frequent hunger, and thirst. Due to these types of factors, diabetes (sugar patients) risk has increased worldwide. The diabetic disease has become a significant issue in the world. This disease is categorized into two types: type 1 and type 2. According to the International Diabetes Federation (IDF), 463 million people worldwide will have diabetes in 2019 and by 2045, this will rise to 700 million. Early detection of diabetic positives helps reduce patients' medical expenditure, death rate, and risk because they may not have proper health care facilities nearby.

The diabetic patient is categorized into two categories, namely type 1 and type 2. Type 1 diabetic patients are dependent on insulin to control the disease. Type 2 diabetic patients are non-dependent on insulin to control the disease. The diabetic-positive patient has a high risk of several problems such as cardiac arrest, kidney failure, dry skin, paralysis, eye problems, etc. Rural area people are unaware of early-stage symptoms to prevent this diabetic disease, and they are unaware of early-stage symptoms to prevent this diabetic disease.

Diabetes patients increase irrespective of age across all regions of the world, and there is no medicine (vaccine) to prevent it. The diabetic positive type 2 patient body makes gradually used insulin. The insulin increases the blood sugar to save energy into the cells for later usage. The diabetic positive diagnosis or confirmation is made at the hospital through conducting various lab/clinical tests. In this modern life, people are interested in saving their time to save money. It leads to many health complications; one of these is diabetic disease. The proposed system is one of the automated processes of early prediction of diabetic positives. There are several machine learning models were proposed for predicting the person with diabetes positively. The proposed model has outperformed in terms of prediction of diabetic positive compared to other existing models and has achieved 93% accuracy as a detection rate.

Limitations on Existing Works

The existing approaches in the prediction of diabetic positive are discussed in the next section. Most of the existing works use only a single algorithm, which is used to predict whether the patient is diabetic positive. There are two problems if one algorithm is used for predicting the output. The first problem is that a single algorithm is not sufficient for prediction. Also, the selected dataset may not fit that algorithm. These problems lead to less accuracy in output prediction. The proposed system has used multiple machine-learning algorithms to predict whether the patient has a diabetic positive or not.

Main Contributions of the Current Work

The proposed system has considered the increasing number of diabetic-positive patients, one of the common problems of all countries globally. Every country has suffered from two main problems without solutions. One is climate change and increasing diabetic positive patients. Thus this paper has investigated the common health issue, diabetic positive, which has no proper software system to predict with high accuracy. This paper has adopted several machine learning algorithms for automating the prediction of diabetic positives. Since the proposed system is a generic model for diabetic prediction, this software system can be used in any region in the world.

LITERATURE REVIEW

Dhomse Kanchan and Mahale Kishor (2) used multiple machine learning algorithms for rare disease prediction. Kavakiotis et al. (3) proposed multiple machine learning models for the diabetic positive prediction. Kononenko (4) surveyed various medical diagnoses using several artificial intelligent approaches. Kandhasamy and Balamurali (5) used various data mining models such as J48, KNN, and Random Forest, SVM, to predict diabetes mellitus under two different situations (one is before pre-processing and another is after pre-processing). Iyer et al. (6) employed two techniques, namely J48 and Naïve Bayes, to classify diabetic patients. The model J48 approach has achieved 74.87% and Naïve Bayes algorithm has obtained 76.96% accuracy in analyzing diabetes. Ashiquzzaman et al. (7) proposed a Deep Neural Network (DNN) to predict the diabetic positive. The DNN model is also adopted to reduce data overfitting.

Yuvaraj and SriPreethaa (8) adopted Hadoop clustering model for prediction of diabetic positive on big data. Sisodia and Sisodia (9) used Decision Tree, SVM, and Naive Bayes to predict diabetes. The SVM has obtained 65.10% of accuracy using SVM, and Naïve Bayes classifier has obtained 73.82%. Negi and Jaiswal (10) developed a machine learning model for diabetic prediction on different global datasets. This approach is a first attempt of diabetic prediction on global datasets. Soltani and Jafarian (11) proposed a Probabilistic Neural Network (PNN) model for diagnosing diabetes type 2 using the PIMA Indians Diabetes data set. This PNN approach has achieved 90% accuracy in analyzing diabetes. Rakshit et al. (12) used a Two-Class Neural Network to predict diabetes. This model has achieved an 83.3% detection rate of type 2 diabetes. Mamuda and Sathasivam (13) compared four machine learning approaches, Naïve Bayes, Quadratic Discriminant Analysis, Linear Discriminant Analysis, and Gaussian Process Classifier, and obtained the accuracy as 81.97% with respect to cross validation of 10.

Farran et al. (14) proposed several prediction models to predict the risk factor of diabetic two positive patients. Anand and Shakti (15) combined multiple machine learning models to predict the diabetic based on personal lifestyle indicators. Malik et al. (16) proposed a non-invasive detection model for blood glucose level using saliva. Mirshahvalad and Zanjani (17) proposed multiple ensemble techniques for diabetes prediction. Mohebbi et al. (18) developed a deep learning model to detect type 2 diabetics. Pham et al. (19) developed a deep learning model for analyzing medical records to predict the trajectories. Askarzadeh and Rezazadeh (20) proposed a neural network model to achieve an effective training novel optimization algorithm for the clinical data analysis. Rao et al. (21) developed a combined classifiers for disease diagnosis. Kopitar et al. (22) employed three techniques, namely Random Forest algorithm, Naïve Bayes classifier, and KNN, for predicting the diabetic. Apart from these machine learning algorithms for predicting

diabetic positive, they also applied XGBOOST, Glmnet, and LightGBM methods for diabetic prediction. Among these methods, the XGBOOST outperformed in diabetic prediction. It has obtained 88% accuracy. Naveen et al. (23) adopted five different machine learning algorithms, SVM, selection Tree, Naive Bayes, Logistic Regression, and KNN, to predict the diabetic positive in the PIMA dataset. This combined machine learning algorithm has obtained 75% accuracy in diabetic prediction.

Butt et al. (24) adopted LSTM to predict the diabetic positive in the PIMA dataset. The machine learning models played a major role in data analysis particularly in clinical data analysis. Thus, the proposed work has adopted machine leaning models for the diabetic positive prediction. Apart from this diabetic positive prediction, the machine learning models are also helpful in other clinical data analysis such as heart disease, cancer tumor, and COVID-19 predictions. The following recent research is evidence for the above statement. Jain et al. (25) proposed several machine learning models to predict COVID-19 positive from B-cell dataset. Shubham et al. (26) proposed deep learning based

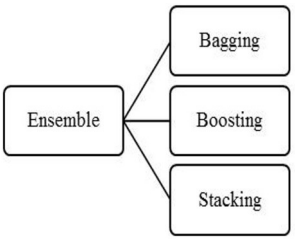


FIGURE 1 | Ensemble techniques.

TABLE 1 | Summary of the existing work.

Author	Proposed	Accuracy(%)	Limitations
Iyer et al. (6)	J48 Naïve Bayes	74.87 76.96	WEKA tool is used for prediction and prediction accuracy rate is less.
Ahmed (26), Singhania et al. (36)	J48 Logistic Regression model Naïve Bayes classifier	73.5 74.4 74.2	A more extensive study is missed for the data analysis.
Soltani and Jafarian (11)	Probabilistic Neural Network (PNN)	89.56	Type 2 diabetics details only considered for the application development.
Kopitar et al. (22)	Naïve Bayes, Random Forest and KNN	64.47	Diabetic prediction accuracy is less compared with proposed stacking approach.
Ashiquzzaman et al. (7)	DNN, with Dropout	88.41	This method is achieved an 88.41 detection rate. Single approach is used.
Chugh et al. (27)	Decision Tree and Gradient Boosting machine	90.00	The proposed method achieved a 90 accuracy in analyzing diabetes. This paper has focused only on children's data for predicting the diabetics.
Rakshit et al. (12)	Two-class neural network	83.3	This proposed model achieved an 83.3 detection rate of type 2 diabetes. This method has considered the women dataset with their age above 21.
Maniruzzaman et al. (13)	Linear Discriminant Analysis, Quadratic Discriminant Analysis, Naïve Bayes classifier, Gaussian Process modeling	81.97	They accuracy as 81.97, which is less than the proposed method.
Sisodia and Sisodia (9)	Decision Tree SVM Naive Bayes	76.30 65.10 73.82	Diabetic prediction accuracy is less compared to proposed stacking approach.
Rao et al. (21)	Decision Tree with radial function	75.65	Diabetic prediction accuracy is less compared with proposed stacking approach.
Kopitar et al. (22)	XGBOOST	88.4	The obtained accuracy is less and single algorithm XGBOOST is used.
Naveen et al. (23)	SVM, selection Tree, Naive Bayes, Logistic Regression and KNN	75	Several algorithms are used but those algorithms are not combined together for final prediction.
Aishwarya et al. (21), Gadekallu et al. (37), Anup et al. (37)	SVM	95	Single machine learning algorithm is used for prediction.
Kandhasamy and Balamurali (5), Meri et al. (38), Ghazal et al. (39, 40), Hasan et al. (41, 42), Siddiqui et al. (43), Upadhyaya et al. (43), Bakri Hassan et al. (44), Salih Ahmed et al. (45), Ahmed et al. (46), Alsharif et al. (47), Khalifa et al. (48)	J48, KNN, RF, and SVM	73.82	Diabetic prediction accuracy is less compared to proposed stacking approach.

for identification of glomeruli in the human kidney. Mohan et al. (27) employed two techniques, namely Decision Tree and Gradient Boosting machine, to predict heart disease. Kumar et al. (28) developed popular RNN model and Reinforcement learning model for COVID-19 prediction. Ngabo et al. (29) proposed several machine learning models and Reinforcement Learning Model for COVID-19 prediction. Iwendi et al. (30) proposed boosted random forest algorithm for COVID-19 disease prediction. Deepa et al. (31) developed an intelligent system based on AI with GDM approaches for healthcare analysis. Dhanamjayulu et al. (32) proposed an image processing technique to identify malnutrition from facial images.

Iwendi et al. (33) proposed a model called N-sanitization, which is used to analyse the unstructured medical datasets for various disease diagnosis. Ahmed et al. (34) used multiple machine-learning models, namely J48, Logistic Regression (LR), and Naïve Bayes (NB). The model of J48 achieved 73.5%, Logistic Regression gained 74.4%, and Naïve Bayes achieved 74.2% with 10-fold cross-validation. Kalra et al. (35) performed a detailed study on diabetic type 1 patients' medical records. **Table 1** shows the accuracy details of diabetic positive prediction with respect to existing works.

METHOD AND TECHNIQUES

Ensemble Techniques

In the conventional approach, only one machine algorithm is used for problem-solving. But the single algorithm is not enough for the complex problems. That algorithm may not fit to the input data due to parameter constraints, input data format constraints, and so on. That is the reason that combining more than two machine algorithms, called an "ensemble model," becomes popular. But the popular question on the ensemble technique is, "How do ensemble models achieve better performance than single approach?" The answer is simple. Just as diversity in nature contributes to more robust biological systems, ensembles of ML models produce stronger results by combining the strengths (and compensating for the weaknesses) of multiple sub models. The proposed system adopts multiple machine learning algorithms (ensemble) to predict the diabetic.

The Ensemble technique has three categories, bagging, boosting, shown in **Figure 1**. Each model has its merit and demerits. Among these three, the proposed system has used stacked ensemble modeling for predicting diabetic positive. **Table 2** shows the performance analysis of three ensemble models. From **Table 2**, the stacking is better compared to the other two models in improving the accuracy. In the healthcare system, prediction accuracy is a significant feature to evaluate the system. Since diabetic positive or negative prediction is under the healthcare system, the stacked ensemble approach is used in the proposed model.

Stacking

Stacking is a two-level classification technique, namely level-0, level-1, or Meta classifier. Unless conventional bagging and boosting, the stacking creates a new training dataset for the final prediction. This approach is entirely different from other multi-classifier algorithms because other multi-classifier approaches use the averaging or voting for the final prediction. But the stacking relays on the predicted probability set, which is generated from all the classifiers. In level 0, more than one algorithm is used. Level 0 works in either a homogeneous or heterogeneous algorithm set. In homogeneous, the same algorithm is used with different parameters, whereas different algorithms are used in heterogeneous. These level-0 algorithms are trained from the original dataset. After the training, the algorithms do not predict the final output. Instead, the probabilities of each class are predicted. Each algorithm predicted the probability of each class and finally generated the predicted probability set. This set will be given as the input to the level-1 algorithm. The level-1 algorithm is trained from the predicted probability set for the final prediction.

TABLE 2 | Ensemble techniques comparisons.

Bagging	Boosting	Stacked
Multiple Classifiers are trained parallelly.	Builds the new learner in a sequential way.	Multiple Classifiers are trained parallelly.
The result is obtained by averaging the responses of the N learners.	On each iteration, update the model by weights until the desired result is obtained.	The result is obtained from the second level classifier.
Reduces the variance.	Reduces the bias.	Increases the accuracy.

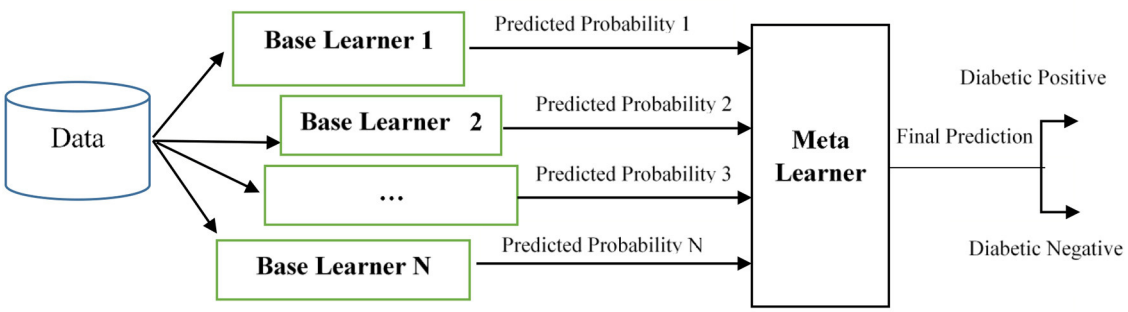


FIGURE 2 | Stacked ensemble model.

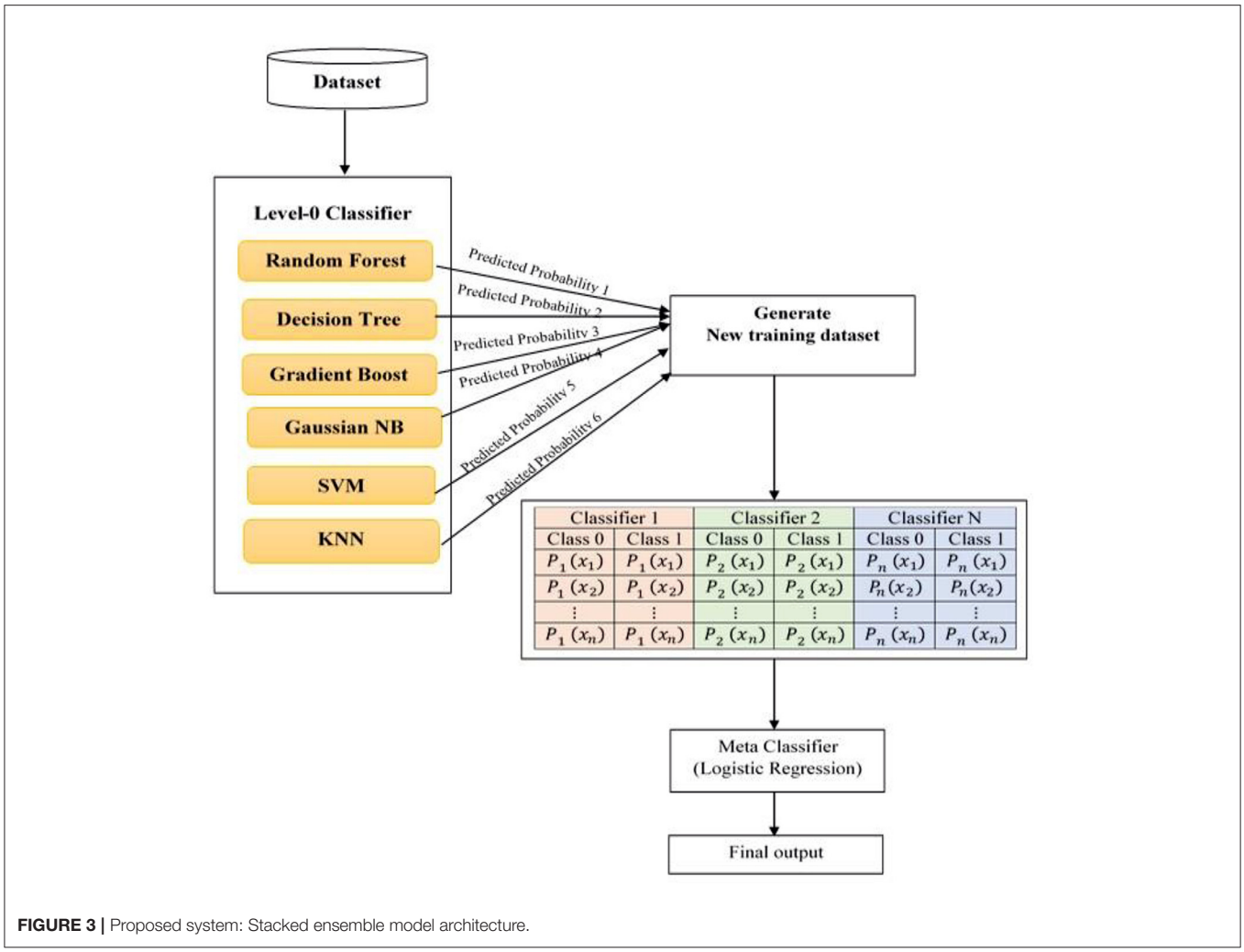


FIGURE 3 | Proposed system: Stacked ensemble model architecture.

The generic model of the proposed stacked ensemble model is shown in **Figure 2**. In **Figure 2**, the base learners 1, 2… N are level-0 classifiers, also called as weak learners. The base learners are trained from the dataset to construct the new training set. Meta learner is the level-1 classifier, and it will be trained with the newly created set. After training, the level-1 classifier will predict the test set.

Level-0 Classifier

Figure 3 depicts the proposed system architecture. The proposed system has selected six different type of machine learning models as level-0 classifiers.

The selected level-0 classifiers are Random Forest, KNN classifier, Decision tree, Gradient Boosting, SVM, and Gaussian Naïve Bayes added as base classifiers. These base classifiers are trained with the original dataset and output a new training dataset for the level-1 classifier.

The advantage of a level-0 classifier in the stacked ensemble technique is two-level classification. In the conventional approach, only one machine algorithm is used for problem-solving. The stacked ensemble learning model is called a multiple classifier system that uses base classifiers to build new training data to classify unknown data. In **Figure 3**, the level-1 classifier logistic regression is represented as a Meta classifier. The level-0 classifier predicted probability output is given as input to the logistic regression Meta classifier.

TABLE 3 | Level-0 Input Set.

Attribute Set	Target output label
$AttrVec_1$ (1st row)	y_1
$AttrVec_2$ (2nd row)	y_2
$AttrVec_3$	y_m
⋮	⋮
$AttrVec_n$ (nth row)	y_2

Level-0 Classifier Input

For the level-0 classifiers, the input is the original dataset. The dataset consists of an input vector set ($AttrVec_1$,

Classifier-1				...				Classifier-k				Actual
a1	a2	...	an	...	...	...	...	a1	a2	...	an	
$P_1(x_1)$	$P_1(x_1)$	...	$P_1(x_1)$	...	...	...	...	$P_k(x_1)$	$P_k(x_1)$	...	$P_k(x_1)$	y_1
$P_1(x_2)$	$P_1(x_2)$	...	$P_1(x_2)$	...	...	...	...	$P_k(x_2)$	$P_k(x_2)$	...	$P_k(x_2)$	y_2
⋮	⋮	⋮	⋮	...	...	...	...	⋮	⋮	⋮	⋮	y_m
$P_1(x_n)$	$P_1(x_n)$	...	$P_1(x_n)$	...	...	...	...	$P_k(x_n)$	$P_k(x_n)$	...	$P_k(x_n)$	y_2

FIGURE 4 | Level-1 classifier input set.

$AttrVec_2 \ldots$ $AttrVec_n$) and the output attribute is ($y_1, y_2, \ldots, y_n$). The format for the level-0 input is given in **Table 3**.

Proposed System Level-0 Classifiers

Random Forest

This is one of the supervised machine learning algorithms. It is used to solve classification and regression problems. The random forest builds the forest from decision trees to solve the problem and improve its performance. The decision tree consists of branches, internal nodes, and leaves. Leaf node represents the final result or class label. Internal nodes are evaluators which decide the branch selection on an attribute (e.g., whether age is eligible to vote or not). The following hyperparameters are used to build the random forest. The number of the estimator is assigned as n_estimators, maximum depth of the tree is assigned as max_depth, minimum number of the split is assigned as min_samples_split, a number of maximum features is assigned as max_features, and the number of a maximum leaf node is assigned as max_leaf_nodes. In addition, n_estimators is a hyperparameter, which indicates the number of decision trees to be generated for the prediction. A higher number of trees is recommended for better prediction, but it may increase the execution time. The default value of n_estimators is 10, and it can be increased up to 500 based on the problem. For the proposed system, the n_estimators is set to 10. That is, every attribute must have min_samples_split samples to divide into two splits. For the proposed method, min_samples_split is set as 2.

Gini index is used to generate the decision tree based on the input dataset. The equation for the Gini index is given below.

$$Gini = 1 - \sum_{i=1}^{c} (pi)^2 \quad (1)$$

C is the total number of classes in the dataset. For our problem, c is assigned as 2. In our Pima dataset, there are a total of eight input attributes and one output class label. The class labels are 1 and 0; 1 indicates the patient has diabetes, and 0 indicates that the patient has not diabetic. *pi* is the probability of selecting the branch among the branches in the ith level

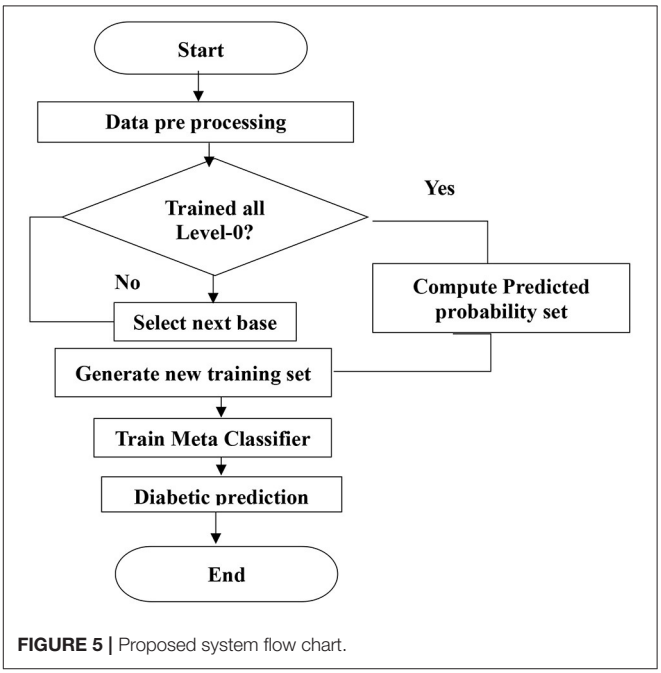


FIGURE 5 | Proposed system flow chart.

for the next level prediction. The proposed system algorithm is shown below.

K Neighbors Classifier

KNN is one of the supervised machine learning algorithms used for classification and regression problems. KNN finds the relationship between the sets X and Y, where X is the input attribute set and Y is the output data. In the KNN, similar training data points are grouped by capturing the distance between the data points. The lesser distance data points are closer than the broader distance data points. Euclidean distance method is used to compute the distance between the data. The following equation is used to calculate Euclidean distance method.

$$d = \sqrt{\sum_{i=1}^{k} (xi - yi)^2} \quad (2)$$

TABLE 4 | Attribute details of Pima Indian Diabetes dataset (PIDD).

SLO.	Feature name with description	Feature name (short)	Variable type	Min val	Max val	Labeled value
1	Number of times pregnant-Number of pregnancy	Pregnant	Integer	0	17	Pregnancies
2	Glucose concentration (2-h oral glucose test [mg/dL])	gl	Integer	0	199	Glucose
3	Blood Pressure (Diastolic blood pressure [mm Hg])	bp	Integer	0	122	Blood pressure
4	Skin thickness (Triceps skin fold thickness [mm])	sk	Integer	0	99	Skin thickness
5	Serum Insulin (2-H serum insulin [mu U/mL])	in	Integer	0	846	Insulin
6	BMI (Body Mass Index [kg/m^2])	bmi	Real	0	67.10	BMI
7	Diabetes Pedigree Function (Diabetes in family history)	dp	Real	0.08	2.42	Diabetes Pedigree Function
8	Age (Age in Years)	age	Integer	21	81	Age
9	Class	Target label	Binary	0 (0-Tested Negative [500])	1 (1-Tested Positive [268])	Target output

ALGORITHM 1 | Algorithm Stacked Ensemble.

1. Input: A training set *D:* = (a_1, *Y*), (a_2, *Y*)… (a_n, Y)
   Input: A testing set T: =(a_1, *Y*), (a_2, *Y*)… (a_m, Y)
   where Y: 0 or 1
   Feature set F: {f_1, f_2, f_3, …, f_n}
2. Step 1: Assign level-0 classifiers
3. Number of level-0 learners *l*=6
4. Step 2: Train the level-0 classifiers using the following
5. **for** *i* = 1 to *n* do
   for *j* = 1 to *l* do
   assign (a_j, b_j) to l_i **Calculate** predicted probability set $\boldsymbol{P}_{li}$
   end for
   end for
6. Step 3: Prepare new training set *(D')*
   D' = **(**$\boldsymbol{P}_{l1}$, $\boldsymbol{P}_{l2}$, $\boldsymbol{P}_{l3}$, **……,** $\boldsymbol{P}_{l6}$, **Y),**
   ($\boldsymbol{P}_{21}$, $\boldsymbol{P}_{22}$, $\boldsymbol{P}_{23}$, **……,** $\boldsymbol{P}_{26}$, **Y),**
   .
   .
   .
   ($\boldsymbol{P}_{n1}$, $\boldsymbol{P}_{n2}$, $\boldsymbol{P}_{n3}$, **……,** $\boldsymbol{P}_{n6}$, **Y)**
7. **Generate Level-0 classifier input set with target output**
   for *i* = 1 to *n* do
8. M_h = (a_1', *Y*), where a_1'=**(**$\boldsymbol{P}_{l1}$, $\boldsymbol{P}_{l2}$, $\boldsymbol{P}_{l3}$, **……,** $\boldsymbol{P}_{l6}$**).**
   M_h – Meta classifier input
9. **end** for
10. Step 4: Assign *(D') to* level-1 classifier (*LR*)
11. Step 5: Train level-1 classifier using *D'*
12. Step 6: **Prepare** testing set (*D"*) for level-0 classifier without target output
13. Step 7: **Execute** level-1-classifier(LR) on *D"*
    for *i* = 1 to *m* do
    ($\boldsymbol{P}_{l1}$, $\boldsymbol{P}_{l2}$, $\boldsymbol{P}_{l3}$, **……,** $\boldsymbol{P}_{l6}$**) predict Y**.
    end for

Decision Tree Classifier

This is a supervised machine learning model that is used for classification. It is a rule-based approach to solve the classification problem. Decision tree is built from the attribute set by applying the if-else pattern set. To create an if-else pattern set or rule set from the attribute, any one of the Gini index, entropy, or misclassification error methods is followed. The most popular approach is the Gini index. These methods are used to create the decision on the internal node and split the samples for the next level in the tree.

$$Gini = 1 - \sum_{i=1}^{c} (pi)^2 \tag{3}$$

$$Entropy = -\sum_{j} p_j \log_2 p_j \tag{4}$$

$$Misclassification\ Error = 1 - max p_j \tag{5}$$

Gradient Boosting

Gradient boosting is a tree-based machine learning algorithm. Boosting is a method that converts weak learners into strong learners. Initially, a tree is built with the dataset attributes and evaluates the model. In evaluation, the error is calculated by original error minus predicted error. This error is also called a classification error. That is, the rate of misclassification is high. This error is minimized or eliminated by building new trees in the subsequent iterations. The error and the first built tree are considered to build the second tree. The second tree is the improved version of the first model, where the misclassification is reduced while compared to the first model. The new tree is built in every iteration using the previous tree classification error

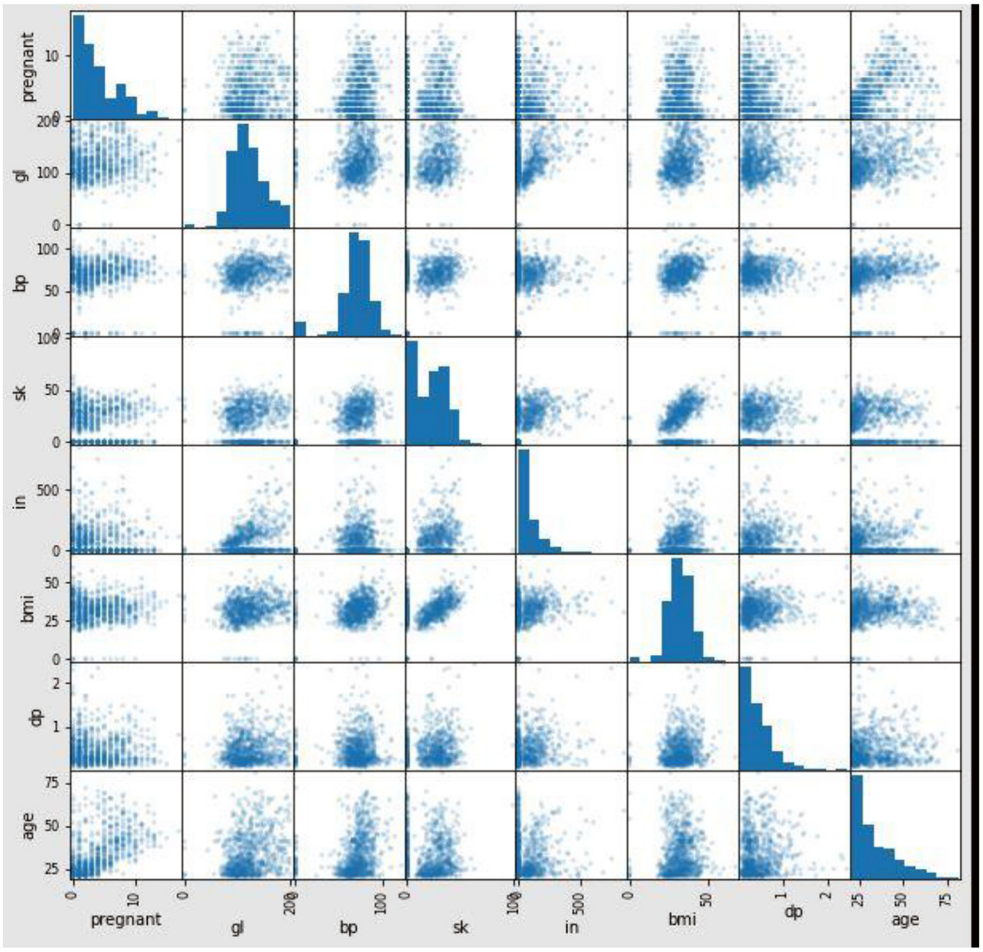


FIGURE 6 | Pearson correlation coefficient of Pima dataset input attributes.

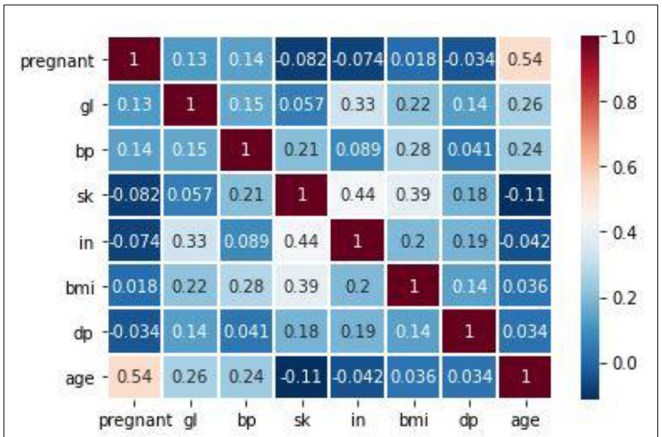


FIGURE 7 | Pearson correlation coefficient result of Pima input attribute set.

and the previous three. This new tree construction is continued until the error becomes negligible or no changes in the error. The following steps are followed for the classification in the gradient boosting approach.

1. Fit a decision tree to the data: $F_1(x)$
2. Fit the following decision tree to the residuals of the previous: $h_1(x) = y - F_1(x)$,
3. Add this new tree to our algorithm: $F_2(x) = F_1(x) + h_1(x)$,
4. Fit the next decision tree to the residuals of F_2: $h_2(x) = y - F_2(x)$,
5. Add this new tree to our algorithm: $F_3(x) = F_2(x) + h_1(x)$,
6. Continue this process until the desired output is reached.

The generic formula of GBM is given in Equation 6.

$$f(x) = \sum_{b=1}^{B} f^b(x) \tag{6}$$

Support Vector Machine (SVM)

SVM is a supervised machine learning approach used for both classification and regression problems. SVM is best suited for classification-related problem-solving approaches. In this approach, the data points are placed in the n-dimensional space, where n is the output classes or features. The SVM is well-suited for binary classification than multiclass

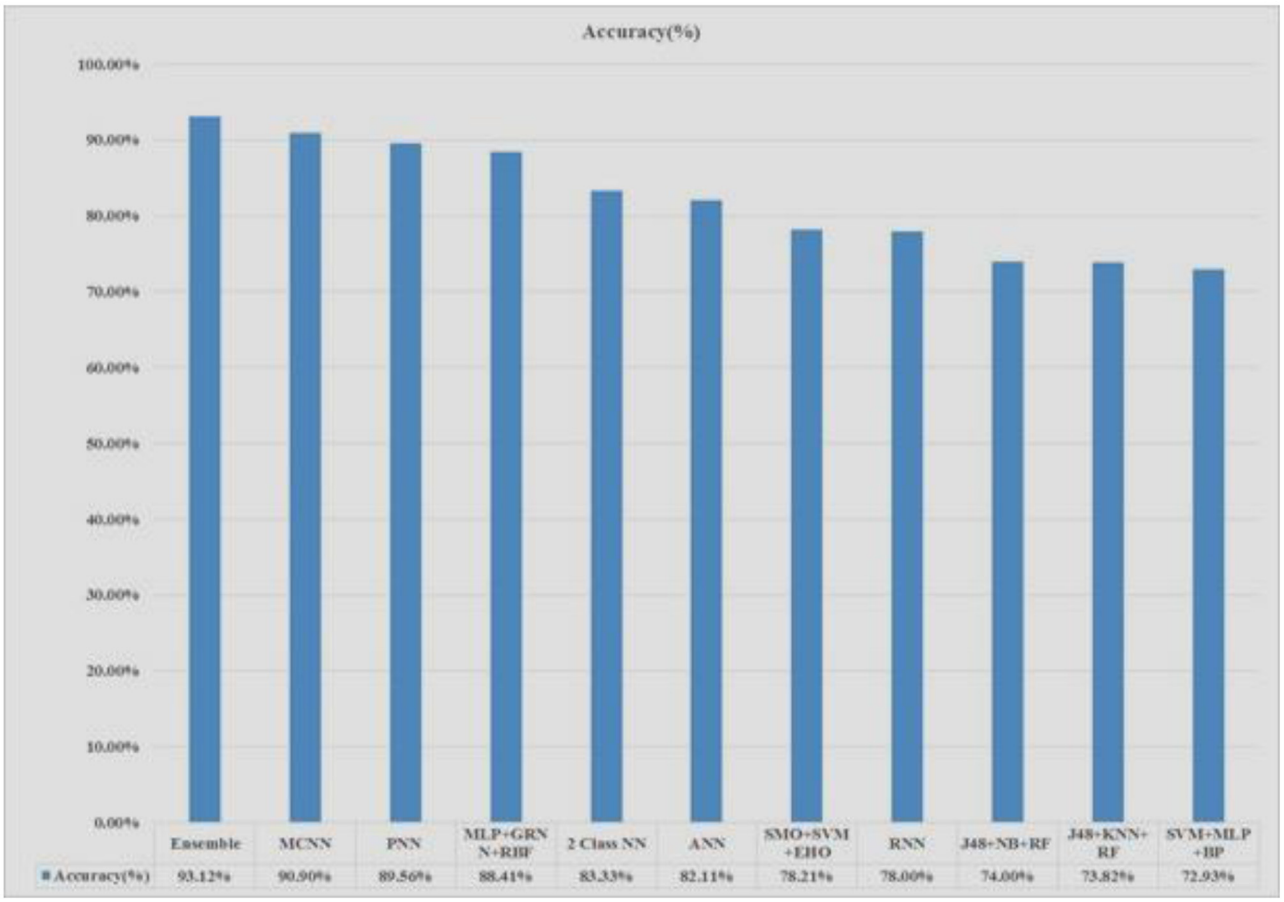


FIGURE 8 | Accuracy chart of various models and proposed model comparison chart.

classification. Equations 7 and 8 are used for computing classification output.

$$w^T X + b \geq c \tag{7}$$

$$w^T X + b < c \tag{8}$$

If the weight matrix and input vector result are higher than c, then the classification output is y1; otherwise, the classification output is y2. Here y1 and y2 are the output class labels.

Gaussian Naïve Bayes

This is a special type of Naïve Bayes approach and suitable for classification problems. It is a supervised machine learning algorithm. It works under the principle of Bayes theorem. The conditional probability calculation is shown in Equation 9.

$$p\,(Y|x_1, x_2, \ldots, x_n) = \frac{p\,(Y)\,p(x_1, x_2, \ldots, x_n|Y)}{p(x_1, x_2, \ldots, x_n)} \tag{9}$$

Level-1 Classifier or Meta Classifier

Figure 4 shows the level-1 classifier input set. For level 1, the logistic regression model is used for the final prediction; this model is trained with a new training dataset generated in level-0 classifiers. The flow chart for the proposed system is shown in **Figure 5**.

This dataset consists of set predicted probabilities of each class of each classifier. A row ri is the predicted probabilities base classifiers of each class of ith row of the original dataset. The formula for the final prediction is done by using Equation 10.

$$p = \frac{e^{b0+b1(x)}}{1 + e^{b0+b1(x)}} \tag{10}$$

$b0$, $b1$ are the constants, and x is the input vector. p is the final prediction, which is >0.5, then the patient has diabetic positive; otherwise, the patient has diabetic negative.

PIMA DATASET DESCRIPTION

Table 4 shows the Pima dataset attributes description. This dataset consists of 768 rows and nine columns. The last column is the output class, containing 1 and 0; 1 indicates diabetic positive, and 0 indicates diabetic negative.

Figure 6 shows the correlation between the attributes in the dataset. The proposed method used Pearson's correlation method, which finds the relationship between the variables in

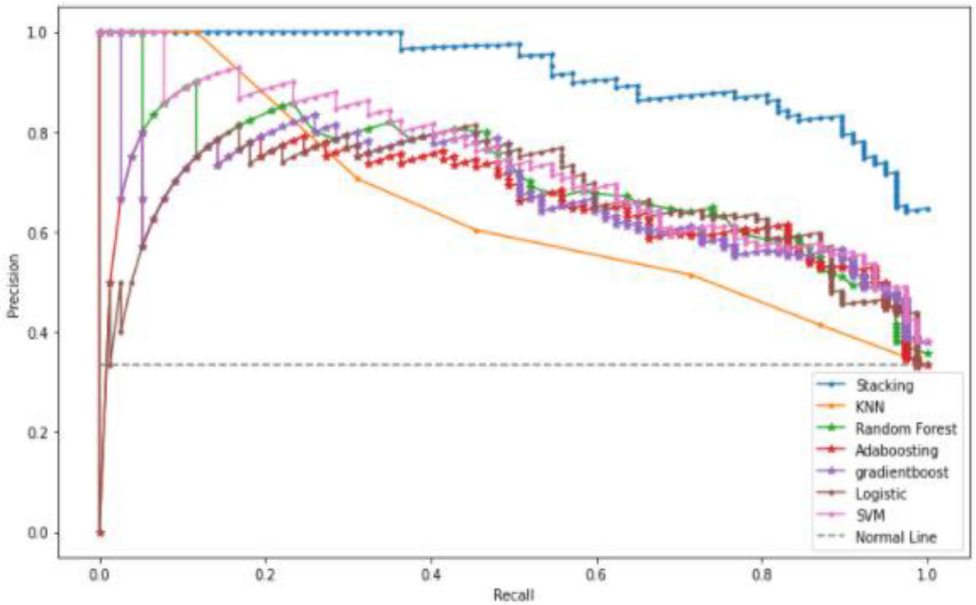


FIGURE 9 | Precision-Recall Curve of the proposed system and various machine learning models.

the Pima dataset. This correlation says how strong an association or correlation of two attributes. Pearson, correlation coefficient formula, returns a value between −1 and 1. The correlation coefficient between two attributes (X, Y) is 1; then, Y's positive value will also increase for every X positive value increase. If the correlation coefficient between two attributes is negative, then any positive value increase of X, Y's negative value will also decrease. If the correlation coefficient is 0, then there is no relation between X and Y. In **Figure 6**, for every attribute pair, the Pima dataset correlation is displayed as a scatterplot. In **Figure 7**, the coefficient value of every two attributes of the Pima dataset is displayed. These two figures depict that most of the attributes in the Pima dataset are independent. The prediction result depends on all the attributes in the dataset.

EXPERIMENTAL RESULTS AND ANALYSIS

The Pima Indians Diabetes (PID) Data Set is applied in the experimental purpose. The whole experiment is done in an Intel Core i5- 6200U CPU @ 2.30GHz 4 cores with 4 Gigabytes of DDR4 RAM with the help of python programming language (32, 33, 37). **Figure 8** shows that the proposed system stacked ensemble model obtained higher accuracy, around 93%, compared to all other existing models.

True Positive (TP) measures correctly predicted the diabetic patients.
True Negative (TN) measures correctly predicted the non-diabetic patients.
False Negative (FN) measures incorrectly predicted the non-diabetic patients.
False Positive (FP) measures incorrectly predicted the diabetic patients.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$$F1 - Score = 2^* \frac{Precision^* Recall}{Precision + Recall} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

A Precision-Recall Curve (PRC) is a metric used to compute the quality of the classifier model. The PRC curve is represented in a graph, where X-axis contains recall values Y-axis contains precision values. This curve depicts the compromise between precision and recall. In a graph, the PRC curve occupies a high area, which means that the obtained recall and precision rates are high. High precision leads to a less false positive rate, and high precision leads to a less false-negative rate. **Figure 9** shows that the proposed stacked ensemble model curve has occupied a higher area than other machine learning models such as KNN, Random Forest, and Gradient Boosting. The curve values are represented as TP/ (TP+FN) on the Y-axis.

The proposed system is compared with other machine learning models by quality metrics such as precision, recall, accuracy, and F1-score. These values are plotted in **Figure 10**. The proposed stacked ensemble model obtained higher results compare to all other methods. **Table 5** shows the quality metrics results. The proposed method is combination of machine

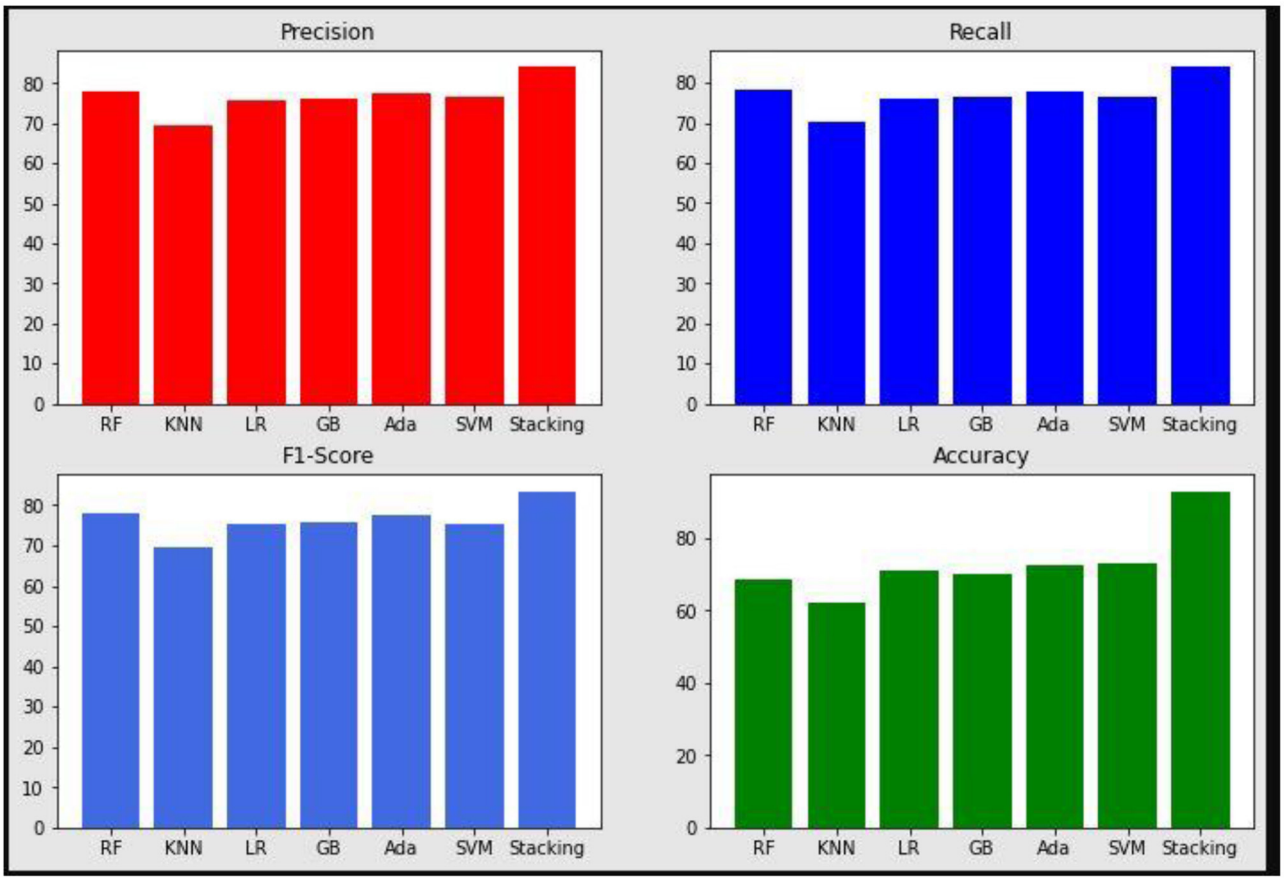


FIGURE 10 | Precision, Recall, F1-Score, and Accuracy results of the proposed system and various machine learning models.

TABLE 5 | Quality metrics results.

Methods	Precision	Recall	F1-Score	Accuracy
Random forest	78	78.3	77.8	68.5
KNN	69.3	70.1	69.5	62.4
Logistic regression	75.7	76.2	75.3	71
Gradient boosting	76.1	76.6	75.9	70
Ada boosting	77.9	77.5	77.9	72.7
SVM	76.5	76.6	75.4	73.1
Stacking	84	83.9	83.5	93.1

learning algorithms. Generally multiple algorithms for a single problem shows better performance. Each machine learning model has its own strengths and weaknesses. If more than one model is combined, then the weakness may be averaged and strength will be increased for many problems, but not all problems. Thus the ensemble techniques such as bagging, boosting, and stacking are popular. Processing time can be higher than single algorithms. The proposed work is also tested with fewer than 6 machine learning approaches with different combination in ensemble technique and obtained lesser than 93% of accuracy of proposed approach.

COMPARED WITH EXISTING WORKS

The proposed stacking ensemble model is compared with other machine learning models. The python language is used to implement the proposed model. And the popular machine learning algorithms such as Random Forest, KNN, Logistic Regression, GBM, etc., are implemented and tested with the PIMA dataset. The obtained result is given in **Table 5**. Compared to the existing models such as Aishwarya et al. (49), Singh et al. (22), and Mamuda and Sathasivam (13), the proposed stacking method has obtained a higher detection rate in detecting the diabetic positive patients.

CONCLUSIONS AND FUTURE WORKS

One of the essential approaches in the medical field is the detection of diseases in the initial stage. Today, diabetes patient increase rates are high irrespective of age across all regions of the world, and there is no medicine (vaccine) to prevent it. The diabetic disease is a big challenge throughout the world, as it affects irrespective of age. Early detection of diabetic positive helps to reduce the medical expenditure,

death rate, and risk of patients. As long as the early prediction on this disease is not famous, the proposed system initiated the prediction of diabetic positive. Experiments are carried out on the Pima Indians Diabetes Database (PIDD). A stacked ensemble model has been adopted in the proposed work and obtained 93% accuracy for a highly categorical dataset.

The existing models in the diabetic prediction used a single algorithm. But the single algorithm will not be suitable for the unstructured and large datasets. Thus, the proposed system has adopted multiple machine learning models called stacked ensemble models. The proposed prediction model has predicted diabetic patients accurately about 93% of the time. In the future, the designed system with the used stacked ensemble method can predict other diseases. The work can be extended and improved for the automation of diabetes analysis, including machine learning and deep learning algorithms.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SR, SM, MH, SA, SA-K, and RS: concept, methodology, validation, analysis, supervision, writing, drafting, review and editing, funding, software, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been supported by the Universiti Kebangsaan Malaysia under research grant GP-2021-K023308, and Taif University Researchers Support, Taif University, Taif, Saudi Arabia.

REFERENCES

1. Muoio DM, Newgard CB. Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nat Rev Mol Cell Biol.* (2008) 9:193–205. doi: 10.1038/nrm2327
2. Dhomse Kanchan B, Mahale Kishor M. Study of machine learning algorithms for special disease prediction using principal of component analysis. *IEEE.* (2016) 5–10. doi: 10.1109/ICGTSPICC.2016.7955260
3. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. *Comput Struct Biotechnol J.* (2017). 15:104–116. doi: 10.1016/j.csbj.2016.12.005
4. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med.* (2001) 23:89–109. doi: 10.1016/S0933-3657(01)00077-X
5. Kandhasamy JP, Balamurali S. Performance analysis of classifier models to predict diabetes mellitus. *Procedia Comput Sci.* (2015) 47:45–51. doi: 10.1016/j.procs.2015.03.182
6. Iyer A, Jeyalatha S, Sumbaly R. Diagnosis of diabetes using classification mining techniques. *arXiv*. (2015) 1–14. doi: 10.5121/ijdkp.2015.5101
7. Ashiquzzaman A, Tushar AK, Islam M, Kim JM. Reduction of overfitting in diabetes prediction using deep learning neural network. *arXiv*. (2017). doi: 10.1007/978-981-10-6451-7_5
8. Yuvaraj N, SriPreethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on hadoop cluster. *Clust Comput.* (2019) 22:1–9. doi: 10.1007/s10586-017-1532-x
9. Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. *Procedia Comput Sci.* (2018) 132:1578–85. doi: 10.1016/j.procs.2018.05.122
10. Negi A, Jaiswal V. A first attempt to develop a diabetes prediction method based on different global datasets. In; *Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. Waknaghat: IEEE (2016). p. 237–41.
11. Soltani Z, Jafarian A. A new artificial neural networks approach for diagnosing diabetes disease type II. *Int J Adv Comput Sci Appl.* (2016) 7:89–94. doi: 10.14569/IJACSA.2016.070611
12. Rakshit S, Manna S, Biswas S, Kundu R, Gupta P, Maitra S, et al. Prediction of diabetes type-II using a two-class neural network. In; *International Conference on Computational Intelligence, Communications, and Business Analytics*. Springer, Singapore (2017). p. 65–71.
13. Mamuda M, Sathasivam S. Predicting the survival of diabetes using neural network. In; *AIP Conference Proceedings*. Kuala Terengganu: AIP Publishing LLC (2017).
14. Farran B, Channanath AM, Behbehani K, Thanaraj TA. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study. *BMJ Open.* (2013) 3:1–10. doi: 10.1136/bmjopen-2012-002457
15. Anand A, Shakti D. Prediction of diabetes based on personal lifestyle indicators. In; *1st International Conference on Next Generation Computing Technologies (NGCT)*. IEEE (2015). p. 673–6.
16. Malik S, Khadgawat R, Anand S, Gupta S. Non-invasive detection of fasting blood glucose level via electrochemical measurement of saliva. *SpringerPlus.* (2016) 5:701. doi: 10.1186/s40064-016-2339-6
17. Mirshahvalad R, Zanjani NA. Diabetes prediction using ensemble perceptron algorithm. In; *9th International Conference on Computational Intelligence and Communication Networks (CICN)*. Girne: IEEE (2017). p. 190–4.
18. Mohebbi A, Aradóttir TB, Johansen AR, Bengtsson H, Fraccaro M, Møru M. A deep learning approach to adherence detection for type 2 diabetics. In; *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jeju: IEEE (2017). p. 2896–9
19. Pham T, Tran T, Phung D, Venkatesh S. Predicting healthcare trajectories from medical records: a deep learning approach. *J Biomed Inform.* (2017) 69:218–229. doi: 10.1016/j.jbi.2017.04.001
20. Askarzadeh A, Rezazadeh A. Artificial neural network training using a new efficient optimization algorithm. *Appl Soft Comput.* (2013) 13:1206–13. doi: 10.1016/j.asoc.2012.10.023
21. Rao NM, Kannan K, Gao XZ, Roy DS. Novel classifiers for intelligent disease diagnosis with multi-objective parameter evolution. *Comput Electr Eng.* (2018) 67:483–96. doi: 10.1016/j.compeleceng.2018.01.039
22. Kopitar L, Kocbek P, Cilar L, Sheikh A, Stiglic G. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Sci Rep.* (2020) 10:11981. doi: 10.1038/s41598-020-68771-z
23. Naveen KG, Rajesh V, Reddy AA, Sumedh K, Reddy TS. Pre-diction of diabetes using machine learning classification algorithms. *Int J Sci Technol Res*. (2020) 9.
24. Butt UM, Letchmunan S, Ali M, Hassan FH, Baqir A, Sherazi HHR. Machine learning based diabetes classification and prediction for healthcare applications. *J Healthc Eng.* (2021) 2021:9930985. doi: 10.1155/2021/9930985
25. Jain N, Jhunthra S, Garg H, Gupta V, Mohan S, Ahmadian A, et al. Prediction modelling of COVID using machine learning methods from B-cell dataset. *Results Phys.* (2021) 21:103813. doi: 10.1016/j.rinp.2021.103813
26. Shubham S, Jain N, Gupta V, Mohan S, Ariffin MM, Ahmadian A. (2021) Identify glomeruli in human kidney tissue images using a deep learning approach. *Soft Comput*. doi: 10.1007/s00500-021-06143-z
27. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access.* (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707

28. Kumar RL, Khan F, Din S, Band SS, Mosavi A, Ibeke E. Recurrent neural network and reinforcement learning model for COVID-19 prediction. *Front Public Health*. (2021) 9:744100. doi: 10.3389/fpubh.2021.744100
29. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng*. (2021) 18:8444–61. doi: 10.3934/mbe.2021418
30. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
31. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using ridge-adaline stochastic gradient descent classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
32. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. IET Image Process. (2021).
33. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-sanitization: a semantic privacy-preserving framework for unstructured medical datasets. *Comput Commun*. (2020) 161:160–71. doi: 10.1016/j.comcom.2020.07.032
34. Ahmed RA, Khalil SN, Al-Qahtani MA. Diabetic retinopathy and the associated risk factors in diabetes type 2 patients in Abha, Saudi Arabia. *J Family Community Med*. (2016) 23:18. doi: 10.4103/2230-8229.172225
35. Kalra S, Das AK, Raghupathy P, Chowdhury S, Saboo BD, Dinakaran P, et al. Current indicators of nutritional care in children with type 1 diabetes in India: do we need a national nutritional guideline? *Indian J Endocrinol Metab*. (2017) 21:670. doi: 10.4103/ijem.IJEM_183_17
36. Singhania U, Tripathy B, Hasan MK, Anumbe NC, Alboaneen D, Ahmed FRA, et al. A predictive and preventive model for onset of Alzheimer's disease. *Front Public Health*. (2021) 9:751536. doi: 10.3389/fpubh.2021.751536
37. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*. (2020) 9:274. doi: 10.3390/electronics9020274
38. Meri A, Hasan MK, Safie N. Success factors affecting the healthcare professionals to utilize cloud computing services. *Asia-Pacific J Inf Technol Multimed*. (2017) 6:31–42. doi: 10.17576/apjitm-2017-0602-04
39. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HMA, et al. Hep-Pred: Hepatitis C Staging Prediction Using Fine Gaussian SVM. *Comput Mater Contin*. (2021) 69:191–203. doi: 10.32604/cmc.2021.015436
40. Ghazal TM, Hasan MK, Alshurideh MT, Alzoubi HM, Ahmad M, Akbar SS, et al. IoT for smart cities: machine learning approaches in smart healthcare—a review. *Future Internet*. (2021) 13:218. doi: 10.3390/fi13080218
41. Hasan MK, Islam S, Sulaiman R, Khan S, Hashim AHA, Habib S, et al. Lightweight encryption technique to enhance medical image security on internet of medical things applications. *IEEE Access*. (2021) 9:47731–42. doi: 10.1109/ACCESS.2021.3061710
42. Hasan MK, Shafiq M, Islam S, Pandey B, Baker El-Ebiary YA, Nafi NS, et al. Lightweight cryptographic algorithms for guessing attack protection in complex internet of things applications. *Complexity*. (2021) 5540296. doi: 10.1155/2021/5540296
43. Siddiqui SY, Haider A, Ghazal TM, Khan MA, Naseer I, Abbas S, et al. IoMT cloud-based intelligent prediction of breast cancer stages empowered with deep learning. *IEEE Access*. (2021) 9:146478–91. doi: 10.1109/ACCESS.2021.3123472
44. Zhao J, Kumar VV. *Handbook of research on innovations and applications of AI, IoT, cognitive technologies*. IGI Global (2021).
45. Ahmed RS, Ahmed ESA, Saeed RA. *Machine Learning in Cyber-Physical Systems in Industry 4.0. In Artificial Intelligence Paradigms for Smart Cyber-Physical Systems*. IGI Global (2021). p. 20–41.
46. Ahmed MM, Hasan MK, Shafiq M, Qays MO, Gadekallu TR, Nebhen J, et al. A peer-to-peer blockchain based interconnected power system. *Energy Rep*. (2021) 7:7890–905. doi: 10.1016/j.egyr.2021.08.071
47. Alsharif S, Saeed RA, Albagory Y. An efficient HAPS cross-layer design to mitigate COVID-19 consequences. *Intell Autom Soft Comput*. (2022) 31:43–59. doi: 10.32604/iasc.2022.019493
48. Khalifa OO, Omar AA, Ahmed MZ, Saeed RA, Hashim A, Esgiar A, et al. *An Automatic Facial Age Prooression Estimation System. 2021 International Congress of Advanced Technology and Engineering (ICOTEN)*. IEEE (2021).
49. Aishwarya R, Gayathri P, Jaisankar N. A method for classification using machine learning technique for diabetes. *Int J Eng Technol*. (2013) 5:2903–8.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SM declared a shared affiliation with the authors to the handling editor at the time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 R., M., Hasan, Saeed, Alsuhibany and Abdel-Khalek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fpubh.2021.825468

An Artificial Intelligence Approach for Expurgating Edible and Non-Edible Items

Dilip Kumar[1], *Urvashi Bansal*[1]*, *Roobaea S. Alroobaea*[2], *Abdullah M. Baqasah*[3] and *Mustapha Hedabou*[4]

[1] Department of Computer Science and Engineering, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India,
[2] Department Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia,
[3] Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia,
[4] School of Computer Science, Mohammed VI Polytechnic University, Ben Guerir, Morocco

In the pandemic of COVID-19, it is crucial to consider the hygiene of the edible and nonedible things as it could be dangerous for our health to consume infected things. Furthermore, everything cannot be boiled before eating as it can destroy fruits and essential minerals and proteins. So, there is a dire need for a smart device that could sanitize edible items. The Germicidal Ultraviolet C (UVC) has proved the capabilities of destroying viruses and pathogens found on the surface of any objects. Although, a few minutes exposure to the UVC can destroy or inactivate the viruses and the pathogens, few doses of UVC light may damage the proteins of edible items and can affect the fruits and vegetables. To this end, we have proposed a novel design of a device that is employed with Artificial Intelligence along with UVC to auto detect the edible items and act accordingly. This causes limited UVC doses to be applied on different items as detected by proposed model according to their permissible limit. Additionally, the device is employed with a smart architecture which leads to consistent distribution of UVC light on the complete surface of the edible items. This results in saving the health as well as nutrition of edible items.

Keywords: food sanitization, smart sanitization, UVC, germicidal, food safety, machine learning for health, smart system

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Zhaoyang Han, University of Aizu, Japan
Mohamed Mounir, El gazeera High Institute for Engineering and Technology, Egypt

***Correspondence:**
Urvashi Bansal
urvashi@nitj.ac.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *30 November 2021*
Accepted: *20 December 2021*
Published: *27 January 2022*

Citation:
Kumar D, Bansal U, Alroobaea RS, Baqasah AM and Hedabou M (2022) An Artificial Intelligence Approach for Expurgating Edible and Non-Edible Items. Front. Public Health 9:825468. doi: 10.3389/fpubh.2021.825468

1. INTRODUCTION

Health is the most critical factor in our life and majorly depends upon what we eat and what is the quality of our daily intake which includes fruits and vegetables having essential minerals, vitamins, and proteins (1). The raw vegetables and fruits may have viruses or pathogens on their surfaces even if some of them are removed by a wash. Due to the COVID-19 pandemic, this becomes essential that the things we are intaking are pathogens free and are virus-free whether it is edible items or our daily usable items. There are many products available in the market for sanitizing edible and non-edible items e.g., UVC rays (2–4). It is already proven that germicidal UVC has properties that deactivate the pathogens by exposing the surface of fruits and vegetables for a few minutes (5–8).

In recent times, it is seen that researchers are widely using UVC to perform complex operations such as scientific applications for deactivating viruses; water purification to remove viruses and bacteria from the water; air purification which deactivates the viruses and germs present in the air of the atmosphere (9); sanitizing a room. Additionally, UVC has many applications for deactivating pathogens over different surfaces and mediums. It is also being used with fluorescent which makes the viruses glow when it is seen by microscopes in presence of the UVC rays. UVC is cheap and easily available in the market which further increases its usage.

In the existing solution, there is a limitation that it damages approximately 30% of the edible items especially fruits and vegetables as its being used by industries for decades for removing the pathogens present on the surface of these items (10–12). So, the existing solution does not have an automated system that can identify the object and make optimal configuration accordingly and an existing solution does not have any filter which prevents the damage done by the exposure of the direct UVC to this item.

We have designed an Artificial Intelligence employed device to provide an automated self-learning process in which the object is identified by the device itself by using machine learning and make optimal configurations as per the class of the object classified by the object detection model. So, AI enables automation (13) with the optimal setting configuration and makes the device smarter than the existing solutions.

2. OUR CONTRIBUTION

In the view of limitations of existing solutions, we have incorporated certain changes in the proposed device so that edible items do not loose its nutrient values.

- We have used an AI-Enabled UVC sanitization device powered by raspberry pi 4, UVC light 254 nm with polarizing filters for the limited dose to the fruits and vegetables, and a camera sensor for object detection.
- We have used the rotating base to apply UVC light uniformly throughout the object.
- The vacuum pump is used to remove dust particles from inside the box which further maintains the uniform intensity of UVC.
- Inside the box, the mirror is presented on each side so total light reflection occurred when the rays from the UVC lamp strikes the mirrors at different angles.
- The aluminum sheet is also sandwiched between the outer shell and the mirror for maximum heat dissipation when the heavy dose is required by the objects.
- Bipolar fans are also present to cool down the aluminum sheets and can dissipate the heat properly out of the box.

We have used machine learning models for automated UVC sanitization. There are many different objects like vegetables and fruits which vary with proteins structure and have different molecules. It is necessary to detect the object placed inside the box so a camera sensor is used to capture the image of the object placed. This object image is passed to the machine learning model which detects the class of the object and then this class is passed to another classifier which classifies the rotation angle and the safe exposure time for that object based on the similarity.

3. LITERATURE REVIEW

UVC Sanitation is being used for many decades in the industry (14). It is not a new thing that is used for sanitizing the different edible and non-edible items. The industry is using germicidal UVC for raw fruits and vegetables to deactivate the pathogens present on the surface of vegetables and fruits (8). Germicidal UVC of 254 nm has the capability of deactivating pathogens as well as air-born virus-like h1n1 flu and also COVID-19 virus (2, 3, 6, 15). Recently, many devices have come into the market which claims that their devices can sanitize any object. There are many products available in the market like air purifiers using the germicidal UVC.

UVC Sanitization box is available in the market in different sizes which can sanitize the items using the germicidal UVC of 254 nm. There is also a UVC device available that can be used for sanitizing small electronic devices and non-electronics devices (5). These devices are good enough for sanitizing the electronic device and the non-edible items but they are not much effective for edible items like raw vegetables and fruits (3).

Some of the findings were provided in **Table 1**. As discussed in existing literature, there is a limitation that a few minutes of doses can lead to damage to edible items like fruits and vegetables. So, we proposed a novel idea to overcome this limitation.

4. METHODOLOGY: SYSTEM ARCHITECTURE OVERVIEW

4.1. Architecture

The newly built proposed design of the AI-enabled UV sanitization box, as shown in **Figures 1**, **2**, depicts the proposed architecture with the following units:

- **Processing unit:** Raspberry pie 4 would be used as a processing unit.
- **Object detection unit:** This is the unit where all the images captured by the image sensor are pre-processed and then the image of the object is passed to the AI trained model which detects the category of object and results in the category of the object. Here, we have created a machine learning model using the CNN and we trained the model with training images of the different fruits and vegetables.
- **Sensors:** We have used a camera module as an image sensor and temperature sensor along with a cooling module.
- **Alarm unit:** We have used speakers and voice enable system to alert the user.
- **Cooling unit:** The same purpose is solved with heat sink and cooling fans.
- **UVC rays filter unit:** UVC 254 nm is very harmful to the human skin as well as it can damage fruits and vegetables. So, to avoid that damage to the vegetables and fruits, the intensity of the light is controlled using the two polarized sheets (23–26). One sheet is fixed and one sheet is made to rotate on both sides directly with the help of a stepper motor. The range of the rotation varies from 0 to 180_0. Intensity can be controlled between 0 and 180_0 from total light pass to total blockage of the light through the polarizing filter (27).
  - **Parallel axis:** When two sheets are placed parallelly, they allowed the total light to passed through it.
  - **In between axis:** If one sheet is fixed and rotation of the sheet is started in either direction, it will start blocking the light as the sheets are getting crossed to each other. If further rotated, it starts allowing the light through it.

TABLE 1 | Existing findings.

Author	Objectives	Finding
Weber et al. (16)	To see how effective Ultraviolet technologies are at reducing microbial contamination on environmental surfaces in patient care environments.	UV sanitization is more likely to be unaffected in darker places.
Scotland (17)	Analyzed the scientific evidence supporting UV light disinfection systems efficiency.	The following factors influenced the effectiveness of UV light: organic load and pathogen; the strength and amount of UV light; the distance from the device; exposure time; surface to be cleaned is within direct line-of-sight or not.
Barbut et al. (18)	The eradication of bacterial spores using hydrogen peroxide sprays and sodium hypochlorite solution was studied in a prospective, randomised, prior to actually trial.	The eradication of bacterial sporesa utilising hydrogen peroxide sprays and sodium hypochlorite solution was studied in a prospective, randomised, before-and-after study. Hydrogen peroxide sprays were shown to be significantly more effective than sodium hypochlorite solution at killing Clostridium difficile spores. The later has the potential to be a viable tool for eradicating Clostridium difficile spores.
Holmdahl et al. (19)	On biological markers, the tests evaluated the effectiveness of hydrogen peroxide and sodium hypochlorite.	On G. Stearothermophilus bioindicators, hydrogen peroxide vapour generators were faster and more effective than sodium hypochlorite machines.
Fu et al. (20)	A comparison of the safety and effectiveness profile of H_2 O_2 sprays against aerosolized hydrogen peroxide on G. Stearothermophilus biological indicators with MRSA, C. Difficile, with Acinetobacterbaumannii discs	The water vapour system has demonstrated a higher level of safety, speed, and efficacy in bacterial inactivation.
Haas et al. (21)	By comparing the frequencies of hospital-acquired MDROs before and after UVD use, a retrospective analysis of the efficiency of UV light environmental decontamination as a complement to improved terminal cleaning of rooms formerly occupied with isolated patients was conducted.	Even though nearly a quarter of the disinfection chances were missed, there was a considerable drop in hospital-acquired MDRO rates during the UVD usage period compared to the previous period. UV technologies appeared to have some positive effects in this investigation.
Anderson (5)	On the surfaces of segregation units, the antibacterial capabilities of UVC light and chemical sanitisers were compared.	UVC proved ineffective in dark portions of the rooms, required further chemical decontamination.
Memarzadeh et al. (22)	The importance of UV light technology in air purification in a healthcare setting is discussed.	UV technologies cannot yet be applied to neutralise or eliminate germs as a stand-alone intervention, but they can be employed in conjunction with other traditional treatments for endpoint disinfection.

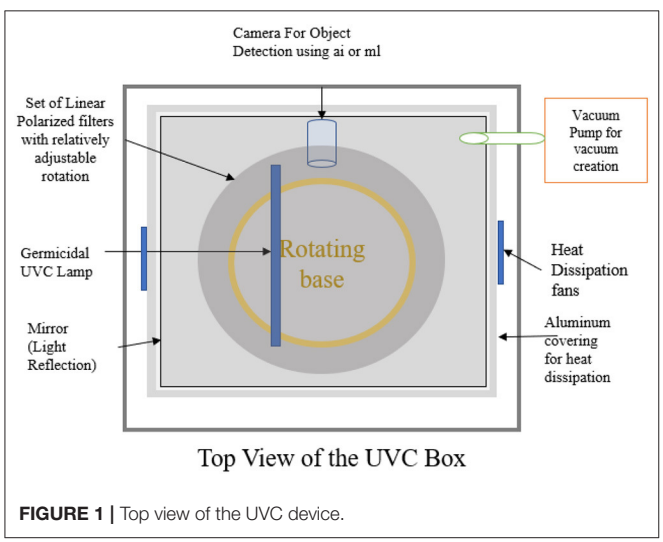


FIGURE 1 | Top view of the UVC device.

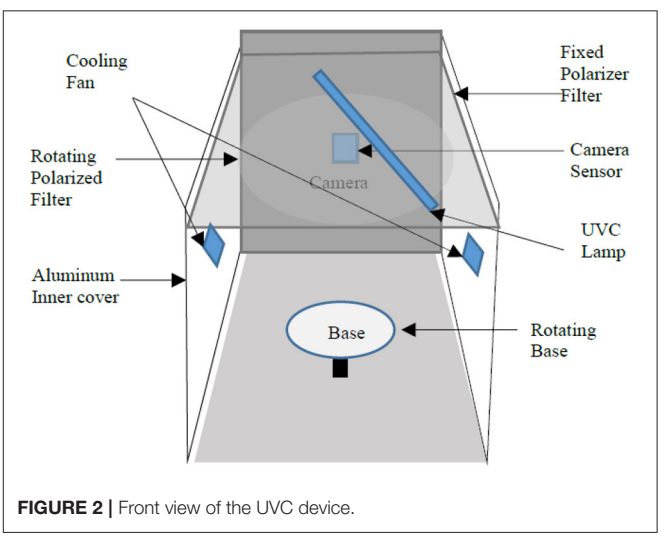


FIGURE 2 | Front view of the UVC device.

- **Crossed axis:** If two sheets are crossed then it blocks the light passing through it.

- **Rotating base unit:** A rotating base is used for the even distribution of the UVC rays to the object which is placed on the rotating base. It also prevents the direct penetration of the UVC. The base starts rotating with the start of the sanitization process.

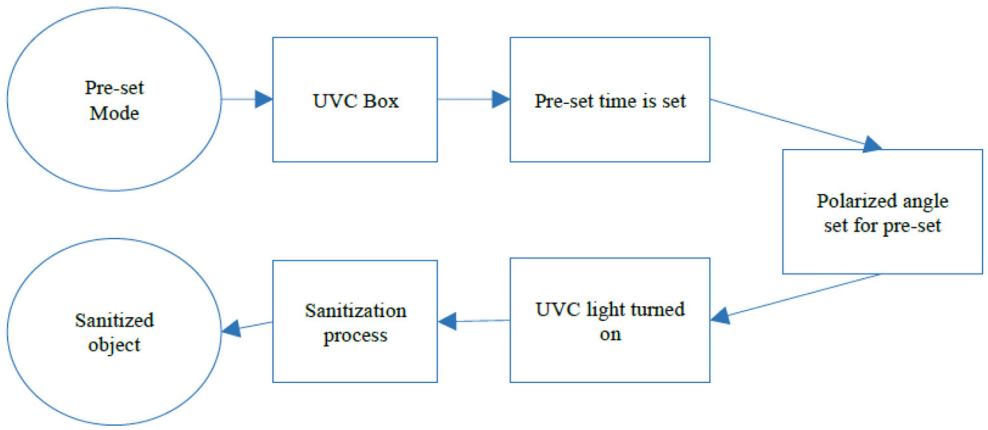


FIGURE 3 | Pre-set mode operation of UVC sanitization.

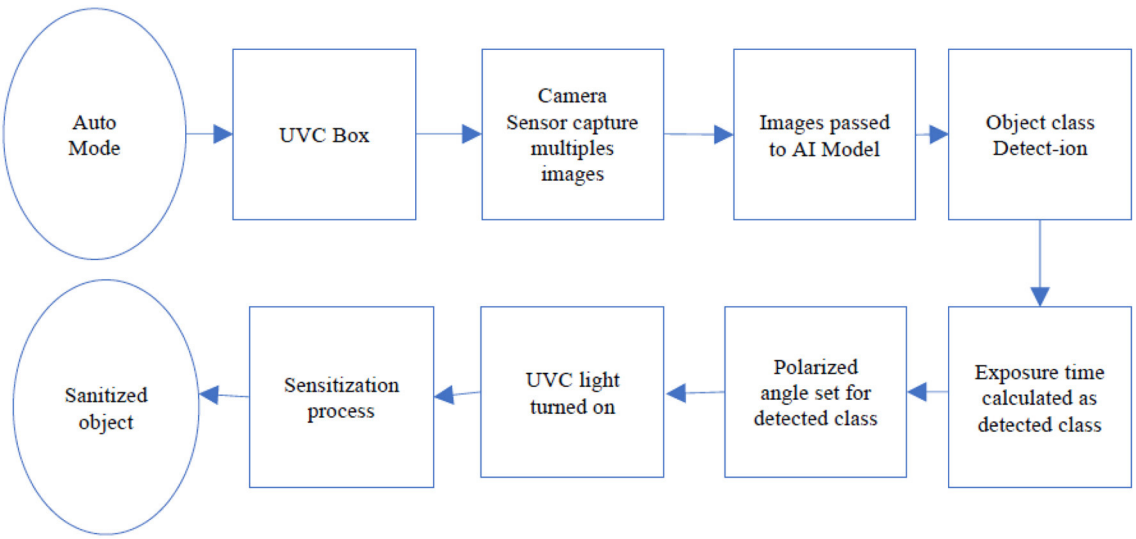


FIGURE 4 | Auto mode operation of UVC sanitization.

- **Display unit:** The display unit displays all events that occurred at the device. It starts from the device start message and mode selected and then displaying the timer set for the UVC sanitization up to the end of the sanitization process.

4.2. Mechanism of Sanitization

Smart UVC Sterilization has a special mechanism of sanitization. It is operated from 230-volt AC. There are two shields: one is the outer shell and one is the inner shell. The opening door consists of an auto cut UVC power supply mechanism which is operated using a switch. It will auto cut the power source to UVC when we open the door and a visible lamp will glow to put the objects inside the UVC box. It has a rotating plate that carries the objects. We have a few pre-set items on the operational menu. We can use those pre-sets if we know the edible item otherwise the auto mode is selected.

When the pre-set mode is selected, the UVC Box is set to the predefined fixed mode where the UVC box works on the predefined values for the UVC exposure times and the UVC box runs for the predefined time and set its polarized angle as defined for that mode (ref **Figure 3**). In the Auto mode, the UVC box uses the AI model to detect the object inside the box. The process is shown in **Figure 4**.

5. CONTROL DESIGN

Control design process of the AI-enabled UVC device consists of different process flows and control structures used in building the device and process description which defines the different sequences of processes held in the device.

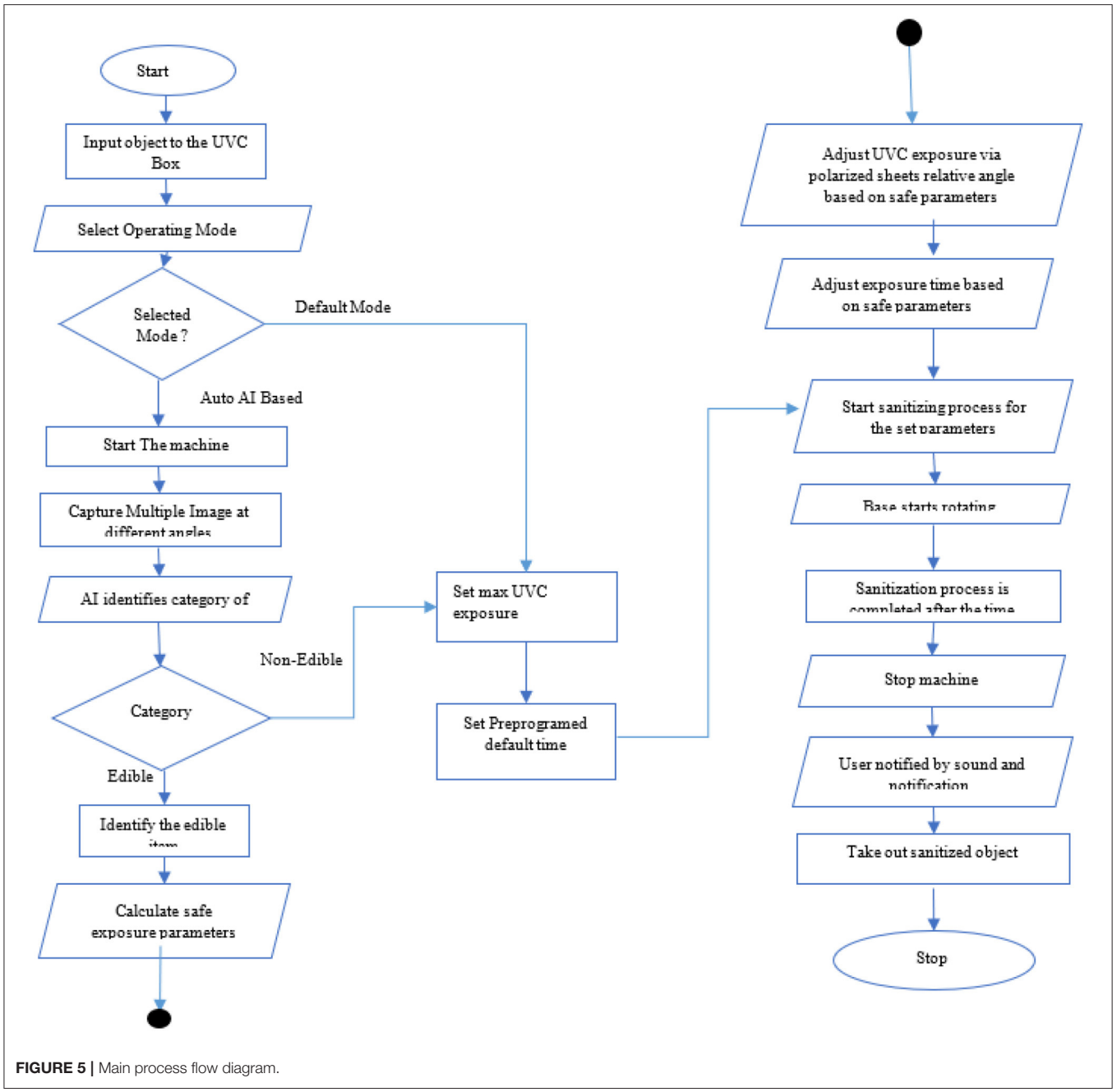


FIGURE 5 | Main process flow diagram.

5.1. Main Process Building Block

This section defines the process of the overall workflow of the AI-Enabled device (28). It defines the overall process of the device from the start to the end of the sanitization process. The main process block is defined in **Figure 5**. The process starts from inputting the edible or non-edible object inside the device and then needs to select the mode of operation of sanitization.

There are two modes: one is manual which can be used directly for the non-edible items; second mode is auto mode which once selected, it starts capturing the images of the object inserted inside the object. After successfully capturing the image, these images are passed to the object detection model which detects the class of the object and the required UVC exposure time is stored using the UVC dosage calculation formula given in equation 1. According to equation 1, we can easily determine that the total UVC dose

TABLE 2 | UV-C doses for SARS-CoV-2.

Viral inactivation (%)	UV-C dose (mJ/cm^2)	Exposure time (s)
90	0.016	0.01
99	0.706	0.32
99.9	6.556	2.98
99.99	31.880	14.49
99.999	108.714	49.42

time is calculated on the basis of the intensity of light dropped for the particular amount of time.

$$UVC_Dose = UVC_rays_Intensity(\mu W/cm^2) \times UVC_Exposure_Time(seconds). \quad (1)$$

On the basis of **Table 2**, we have calculated the expected doses for the predicted item class. If the class of the object is detected to edible items, then the safe exposure parameters are set according to the detected object, and the rotation angle of the polarizing filter and time of UVC rays exposure is measured and stored using another classifier that classifies the edible item based on the few parameters. Finally, the sanitization process gets started. The procedure is provided in **Algorithm 1**.

5.2. Control Process Flow of Camera Sensor

The Camera Sensor is used for capturing the images of the inserted object whenever the auto mode is selected. AI comes into the picture when auto mode is selected and there is a need to classify the class of the object inserted inside the device. Image captured by the camera sensor is saved to the raspberry pi and then it is transferred to the machine learning model which is already trained over the 131 classes of fruits and vegetables. The camera sensor helps to detect the class of the object inserted inside the device.

5.3. Object Detection Process Flow

The object detection process flow is the process that defines the classification of the input Image *via* a pre-trained model loaded to the raspberry pi. The object detection model is trained over 65,000 images with 131 classes. Image is passed to the model and the model predicts the class and returns the class of the object if it is predicted otherwise return nonedible.

5.4. Polarizer Filter Process Flow

The Polarizer filter process depicts the process of principle used for controlling the intensity of the UVC rays *via* changing the angle between two polarized sheets. One sheet is mounted with the stepper motor which rotates the motor in both directions. When two sheets are crossed to each other it is 180 degrees angle and zero for full light pass through the polarized sheet. The process starts with the configuration settings from the detected object class and starts rotating to the set angle for the detected class.

Algorithm 1| UVC_Process.

```
Intialise: object_class= "edible", mode=0, uvc_time=0, filter_angle=0
Select the mode of operation and set mode equal to mode selected.
if mode = manual then
    go to step 8
else if mode=auto then
    go to the step 15
end if
Set the uvc_time to 300 s
Set the filter_angle equal to 0.
Rotate the polarizer sheet to angle equal to filter_angle.
Turn on UVC Lamp.
Start Rotating Base.
Run the Sanitization process for the set uvc_time.
go to step 21.
Capture an image of the input item.
Detect the class of the object using AI model and set object_class equal to detected object class.
Detect the filter angle and optimized UVC sanitization time using optimized setting_finder_model.
Set the uvc_time equal to detected optimized detected time
Set the filter_angle equal to the detected polaroid angle of rotation.
Rotate the filter upto the filter_angle.
Start the sanitization process for the set time equal to the uvc_time.
Stop UVC light on completion of the set time.
Rotate the polarizer back to its original position.
Stop Rotating base.
```

5.5. Cooling System Process Flow

The cooling system is an essential part of the device as it takes care of the heat generated inside the device and prevents the excessive heat generated by the UVC rays. To avoid damage from the heat generated by the UVC rays, the cooling system helps the aluminum covering to act as the heat dissipation heat sink (26, 29).

The process of the cooling system consists of a heat sensor module where the auto cut relay is programmed with a temperature setting. The cooling fans control the temperature of the device. The temperature is set to the desired value. The threshold value for the temperature is set to 30^0C. Whenever the temperature reaches above 30_0C, it turns on all fans connected to the temperature module. If the temperature goes below 30^0C, it stops all the fans. This process continues whenever there is a temperature rise and down.

6. CIRCUIT DESIGN

The Circuit design includes the interfacing of the different hardwares. This section depicts the interfacing of the different hardwares used in this device.

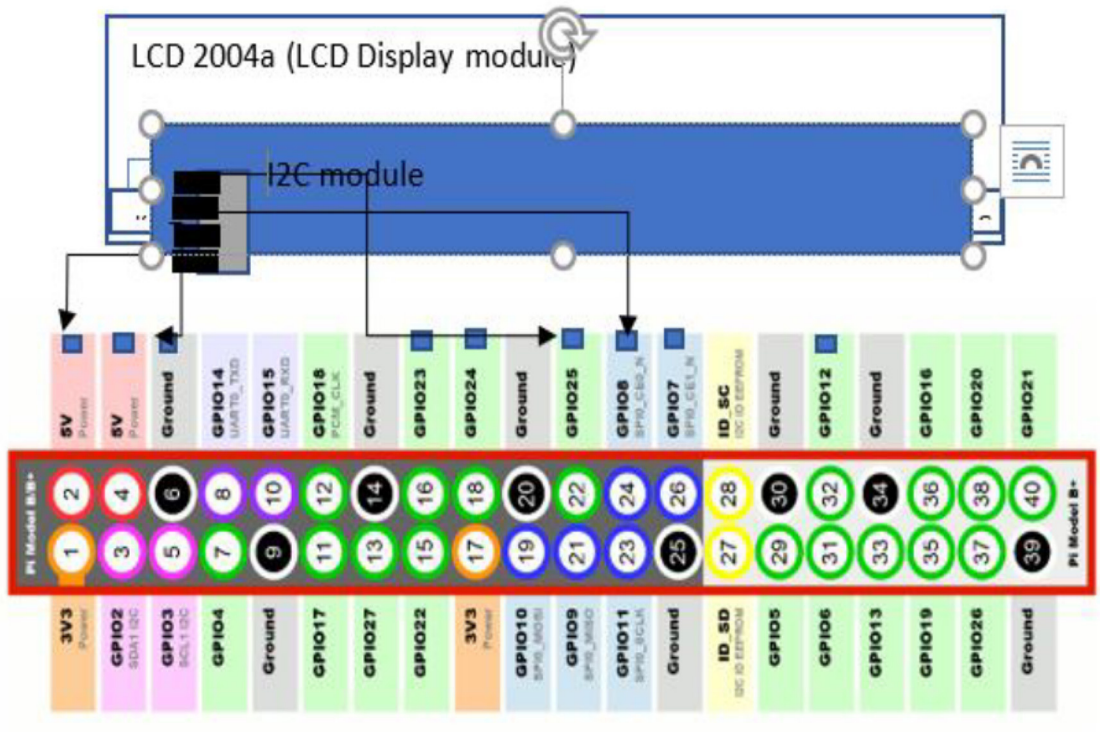


FIGURE 6 | Interfacing of 2004a display LCD with Raspberry Pi.

6.1. 2004a LCD Display Interfacing

LCD 2004 is a parallel LCD display which is simple and a cost effective solution for adding display characters on a 20×4 white and high contrast white text upon a blue background/backlight. The display has 20 characters and 4 lines. It requires 6 data pins to be interfaced with the raspberry pi GPIO pins and other 6 pins with the ground and 5 volts.

The 2004a LCD is interfaced using the i2c adapter which reduces the connection wires from 16 to only a two-wire connection to the I2C adapter. I2C adapter supports many connections to be handled at a time as defined in **Figure 6**. So, it is used here to reduce the complexity of the raspberry pi. It reduces the number of GPIO pins used by the LCD.

6.2. NEMA 17 Stepper Motor Interfacing

The Nema17 stepper motor is used for rotating the object holding base. It has high torque which can rotate weight up to 4 kgs. The L298N stepper driver is interfaced with the Nema17 stepper motor and with a raspberry pi to drive the NEMA as shown in **Figure 7**. Stepper motor drivers need to be powered with 12 volts to drive the NEMA motor.

6.3. RJ2003 Stepper Motor Interfacing

It is a stepper motor that consists of 4 coils which allow it to take very precise steps and very accurately it can rotate up to specific angles rotations. It is used for the rotation of the polarizer filter. It makes the filter from a 0-degree angle to 180-degree angle, which means from a total block of light to the total pass of the light through the polarizer filter. It is interfaced with the raspberry pi. In **Figure 8**, a stepper driver is used to drive the stepper motor that is ULN2003 which is interfaced with both the raspberry pi and stepper motor 28BY-J-48y.

7. PROPOSED TECHNIQUE

7.1. Proposed Object Detection Model

We have used the deep learning method to train the model against the fruits and vegetable dataset of images. We obtained the dataset which is freely available named Fruits-360 which can be downloaded from GitHub or available at Kaggle. It includes about 130,000 images of different categories. This dataset includes high-definition images which are essential to obtain a good classifier (30–33).

The structure of the model we used is defined in **Figure 9**. There are a total of 11 layers used in our model. There are four convolutional layers, 4max pooling layers and 2 fully connected layers, and the output layer is the SoftMax layer. The input layer is a convolutional layer where the filter of 16, 5 x 5 x 4 filter is applied and the max-pooling of 2 x 2 is used with stride=2. Convolutional layer 4 is with parameter (5 x 5 x 64) with dimension 128. The model is trained with the 131 classes of fruits

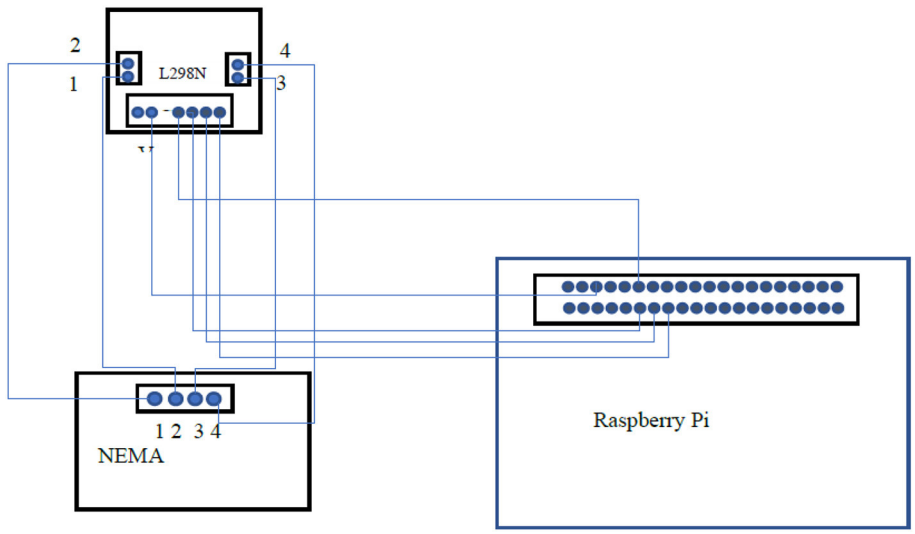


FIGURE 7 | Interfacing of NEMA 17 stepper motor with Raspberry Pi.

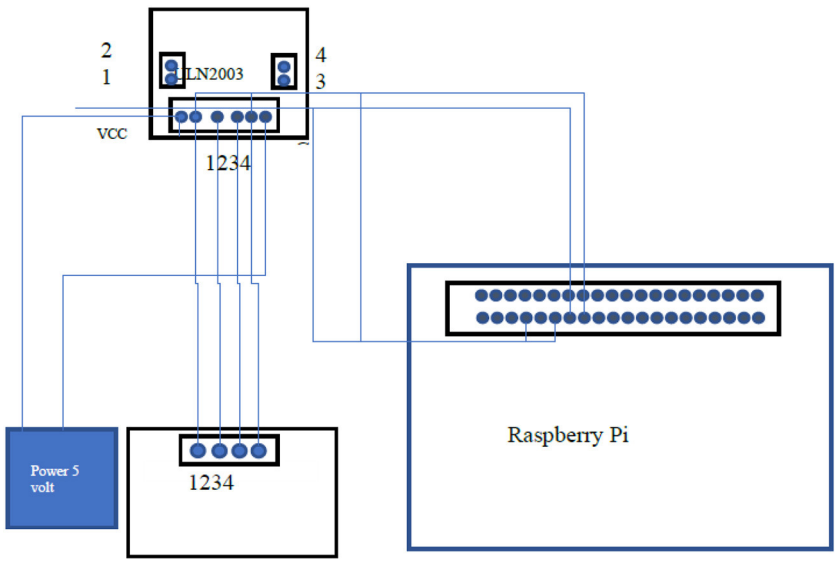


FIGURE 8 | Interfacing of RJ2003 stepper motor with Raspberry Pi.

and vegetables and it can be capable of detecting the multiclass of the image also.

7.2. Optimal Configuration Prediction Model

This is the model proposed to predict the optimal configuration for the detected object. This model makes a prediction based on the data collected through the experimentations. This collected dataset includes the entity name, its features and configuration such as time of UVC exposure and the angle of rotation of the polarizer.

The model used is a decision tree classifier and it works fine as per the requirement. We can classify the UVC exposure time and the angle of rotation of the polarizer and be able to detect the required configuration of the UVC sanitization process. This classifier gives an accuracy score of the 96.33%.

7.3. Smart Device Control

The device is controllable through a mobile app (android). There are two interface screens of the app. The use of the mobile app is to control the device using the Bluetooth communication channel and can be operatable using the mobile app interfaces (34–40). The wireless channel enables the message transfer from mobile to operate the device using the host program running on the raspberry pi and client program on the android app.

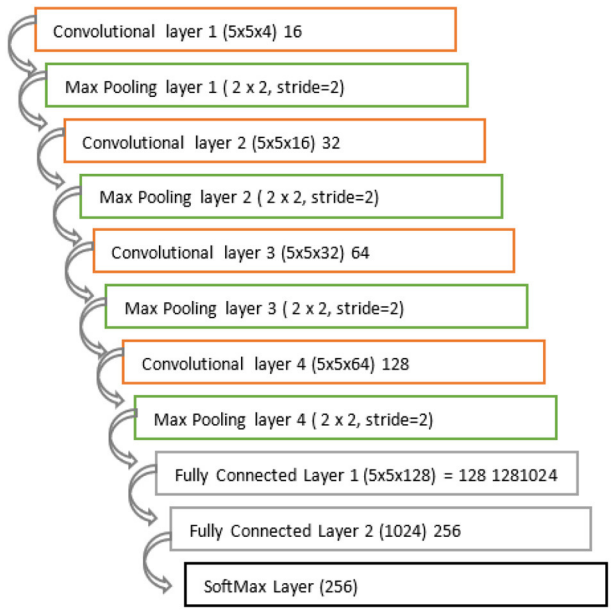


FIGURE 9 | CNN model layers.

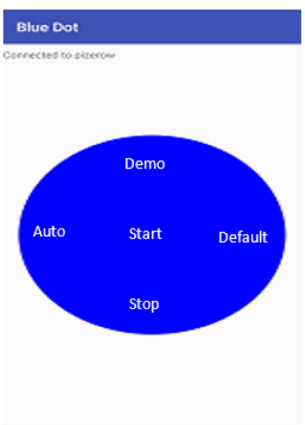


FIGURE 10 | Mobile app main screen with controls.

7.4. Control Screen

The device is configured for the following operations:

7.4.1. Mode Selection

There are two select modes

- Auto mode
- Default mode

The two modes can be selected from app as shown in the app screen in **Figure 10**. One of the mode is the auto mode where the AI comes into the picture and takes care of the rest of the process of the sanitization. The other mode is the default mode operating normally for non-edible items. Its configuration is fixed and loaded on the selection.

8. RESULTS AND DISCUSSION

The code is implemented two times, one experiment is done with 20 items with 20 iterations, another one is done with the 25 items with 20 iterations. The goal is to create an optimal configuration for the discovered items.

8.1. Data Set

In this research realm, the two data sets taken are listed below:

- Fruits 360: This dataset includes the 131 classes of images of both fruits and vegetables (41).
- Configuration mapping: This dataset is made from the experiments performed by us for optimal configuration for the different objects.

8.2. System Requirements

The code of the proposed object detection and training model is written in Python and TensorFlow libraries. The code has been implemented to detect the object class and make optimal configuration using the classification model on experiments results to assess the performance of the UVC process algorithm. The following specifications are used to implement the algorithm:

- Processor - Intel(R) i7-8750H CPU @ 2.40GHz 2.40 GHz
- GPU - NVIDIA-GeForce GTX-1050
- RAM - 8 GB DDR4
- Hard disk drive - 256 SSD
- Operating System - Windows 10 Pro x64-bit

9. RESULTS

The machine learning model used is CNN for object detection and we have used the fruits360 dataset (41) and 50 epochs for training the model which gives an accuracy of 96.2% without overfitting and underfitting as we have provided enough training data for training the model. We have trained the model on an i7 intel processor with Nvidia graphics as it is not possible to train a model over the raspberry pi. Due to limited hardware configuration, we trained it over the system and used the trained model in raspberry pi. So, the trained model gives a good result on the raspberry pi as it is trained over 131 classes of the fruits and vegetables of 65,000 images. We have run the device for the different configurations and the different filter parameters. We have run the machine for the following angle of filters as follows:

- 0^0 angle filter rotation: Here, the light passes through the filters as the normal light passes as it permits the whole rays to pass through the filters.
- 45^0 angle filter rotation: This angle of rotation allows lights to pass through the filter but with limitations as it does not allow the whole UVC light to get pass through it. It allows 75% of UVC light through.
- 90^0 angle filter rotation: At this angle of rotation of the filter, 50% of rays are blocked and 50% of rays get passed through the filters.

TABLE 3 | Experimental results.

Item	45^0 Angle time(s)	Remark	90^0 Angle time(s)	Remark	135^0 Angle time(s)	Remark
Apple	120	Ok	180	Ok	300	Ok
Banana	120	Ok	180	Ok	300	Ok
Grapes	120	Ok	180	damage	300	damage
Papaya	110	Ok	175	Ok	280	Ok
Kiwi	123	Ok	185	Ok	290	Ok

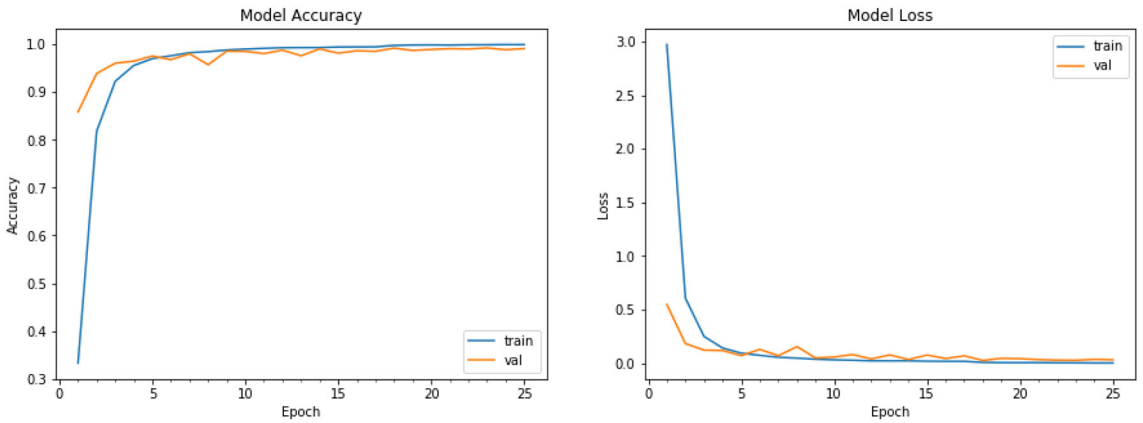


FIGURE 11 | Training accuracy vs. validation accuracy (Left) training loss vs. validation loss (Right) performance graph.

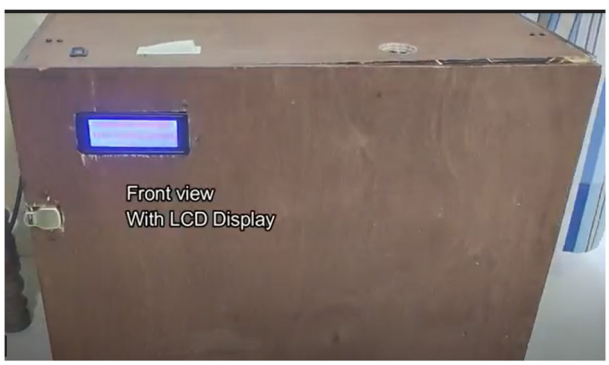


FIGURE 12 | Device front view.

FIGURE 13 | Device internal view.

- 135^0 angle filter rotation: At this angle of rotation of the filter, 70% of rays are blocked and 25% of rays get passed through the filters.
- 180^0 angle filter rotation: At this angle of rotation of the filter, 100% rays are blocked and 0% rays get passed through the filters.

So, we have used the 45-degree, 90-degree, and 135-degree rotations for our experiments and get the following results for a few of the fruits as shown in **Table 3**. We have calculated the UVC dose time using equation 1. As our UVC lamp has a UVC intensity of 300 $\mu W/cm^2$ and the length we have to use is 15 cm.

As represented in **Table 3**, the edible items (especially raw fruits and vegetables) can be sanitized without losing their quality and without any damage to the cell body of the edible item using the proposed system architecture.

9.1. Performance Plot

We have plotted the performance graph for our model for the training data and the validation data. We plot the graph between training data loss vs. validation data loss and training data accuracy and validation data accuracy (**Figure 11**) of the CNN model for object detection. The performance graph plot tells the increase in accuracy with the increase in the epochs and it also describes whether the model is trained properly or not. **Figure 11**

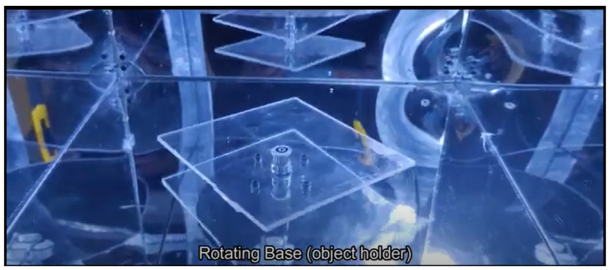


FIGURE 14 | Device internal view - rotating base.

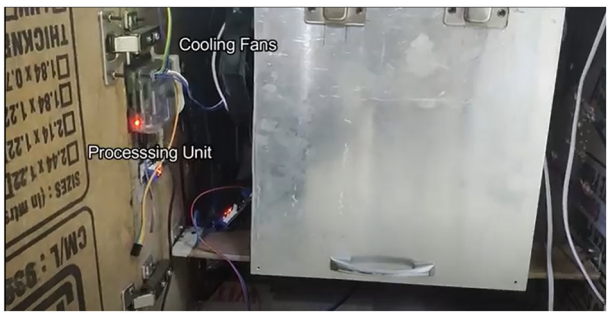


FIGURE 15 | Device internal view - processing unit and cooling fans.

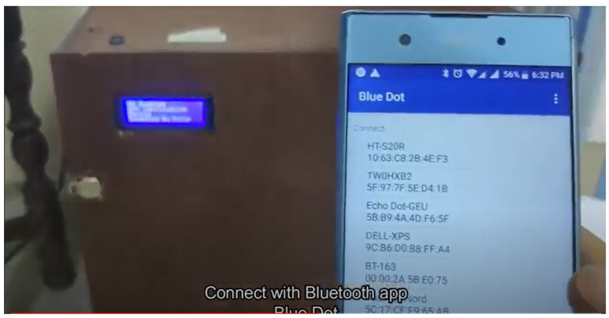


FIGURE 16 | Device connectivity.

is the graph plotted between the loss at training and loss at Validation time. As portrayed in the graph, it is nearly converging so there is no overfitting and underfitting in our model. Training accuracy vs. validation accuracy performance graph shows that the model is trained nearly perfect as there is very little difference between the validation and the training accuracy. So, our model gives an accuracy of the 96.2% which is good enough to detect the type of vegetable and fruits from an image and it works fine for the single item or single fruit or vegetable image. It is also capable of predicting the class of multiple objects too. The model works fine for 131 types of categories. It gives very accurate results for the images input to the model. Finally, **Figures 12–16** provide some additional images of the device under testing. These images are for illustration purposes only.

10. CONCLUSION AND FUTURE SCOPE

With the proposed system architecture, the edible items (especially raw fruits and vegetables) can be sanitized without losing their quality and without any damage to the cell body of the edible item. The AI-enabled device sanitizes the edible items with different optimal configurations made using object detection and automated configuration detection based on the detected class. The optimal configurations utilize the phenomenon of polarization filters which limits the expose of rays due to which the edible items get sanitized without losing their nutrition. Limited filter UVC doses are exposed uniformly throughout the surface of the item with different configurations with the optimal time.

In future, some points need to be considered to enhance the research work. The developed prototype is smart enough to detect the object inserted inside the box but the UVC 254 nm is not safe for humans as well as for edible items. So, 222 nm Germicidal UVC light can be used instead of 254 nm if it is available. Smart Home integration and API could be done to control the device using IoT, e.g., google assistant. Simultaneous sanitization could be performed for multiple items using optimal parameters.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.kaggle.com/moltean/fruits.

AUTHOR CONTRIBUTIONS

UB and DK carried out the experiment. DK wrote the manuscript with support from UB. DK fabricated the device with the help from RA, AB, and MH. UB supervised the project. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors are grateful to the Taif University Researchers Supporting Project (Number TURSP-2020/36), Taif University, Taif, Saudi Arabia and Dr. B. R. Ambedkar National Institute of Technology Jalandhar for their support to carry out this work.

REFERENCES

1. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with Fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inform*. (2021) 1–6. doi: 10.1109/JBHI.2021.3112693
2. Santhosh R, Yadav S. Low cost multipurpose UV-C sterilizer box for protection against COVID-19. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. Coimbatore: IEEE (2021). p. 1495–8.
3. Zaffina S, Camisa V, Lembo M, Vinci MR, Tucci MG, Borra M, et al. Accidental exposure to UV radiation produced by germicidal lamp:

case report and risk assessment. *Photochem Photobiol*. (2012) 88:1001–4. doi: 10.1111/j.1751-1097.2012.01151.x
4. Guerrero-Beltr·n J, Barbosa-C·novas G. Advantages and limitations on processing foods by UV light. *Food Sci Technol Int*. (2004) 10:137–47. doi: 10.1177/1082013204044359
5. Anderson M. The ultraviolet offense: Germicidal UV lamps destroy vicious viruses. New tech might put them many more places without harming humans. *IEEE Spectrum*. (2020) 57:50–5. doi: 10.1109/MSPEC.2020.9205549
6. Nenova V, Nenova MV, Georgieva DV, Gueorguiev VE. ISO standard implementation impact in COVID-19 erra on UV-lighting devices. In: *2020 Fifth Junior Conference on Lighting (Lighting)*. Ruse: IEEE (2020). p. 1–4.
7. Cao Y, Chen W, Li M, Xu B, Fan J, Zhang G. Simulation based design of deep ultraviolet LED array module used in virus disinfection. In: *2020 21st International Conference on Electronic Packaging Technology (ICEPT)*. Guangzhou: IEEE (2020). p. 1–4.
8. Yamano N, Kunisada M, Kaidzu S, Sugihara K, Nishiaki-Sawada A, Ohashi H, et al. Long-term effects of 222-nm ultraviolet radiation C sterilizing lamps on mice susceptible to ultraviolet radiation. *Photochem Photobiol*. (2020) 96:853–62. doi: 10.1111/php.13269
9. Alyami H, Alosaimi W, Krichen M, Alroobaea R. monitoring social distancing using artificial intelligence for fighting COVID-19 virus spread. *Int J Open Source Softw Process*. (2021) 12:48–63. doi: 10.4018/IJOSSP.2021070104
10. Dhanamjayulu C, Nizhal U, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) 1–12. Available online at: https://doi.org/10.1049/ipr2.12222
11. Mukhtar H, Rubaiee S, Krichen M, Alroobaea R. An IoT framework for screening of COVID-19 using real-time data from wearable sensors. *Int J Environ Res Public Health*. (2021) 18:4022. doi: 10.3390/ijerph18084022
12. Arul R, Alroobaea R, Tariq U, Almulihi AH, Alharithi FS, Shoaib U. IoT-enabled healthcare systems using block chain-dependent adaptable services. *Pers Ubiquitous Comput*. (2021) p. 1–15. doi: 10.1007/s00779-021-01584-7
13. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Intern Things J*. (2021) 14:1–9. doi: 10.1109/JIOT.2021.3117762
14. Bourouis S, Channoufi I, Alroobaea R, Rubaiee S, Andejany M, Bouguila N. Color object segmentation and tracking using flexible statistical model and level-set. *Multimed Tools Appl*. (2021) 80:5809–31. doi: 10.1007/s11042-020-09809-2
15. Juarez-Leon FA, Soriano-Sánchez AG, Rodríguez-Licea MA, Perez-Pinal FJ. Design and implementation of a germicidal UVC-LED lamp. *IEEE Access*. (2020) 8:196951–62. doi: 10.1109/ACCESS.2020.3034436
16. Weber DJ, Rutala WA, Anderson DJ, Chen LF, Sickbert-Bennett EE, Boyce JM. Effectiveness of ultraviolet devices and hydrogen peroxide systems for terminal room decontamination: focus on clinical trials. *Am J Infect Control*. (2016) 44:e77–e84. doi: 10.1016/j.ajic.2015.11.015
17. Scotland HP. Literature review and practice recommendations: existing and emerging technologies used for decontamination of the healthcare environment: hydrogen peroxide. *Natl Serv Scotland*. (2016). Published on December 5 2016, Available online at: https://www.nss.nhs.scot/publications/literature-review-and-practice-recommendations-existing-and-emerging-technologies-for-decontamination-of-the-health-and-care-environment-airborne-hydrogen-peroxide-v1-1/
18. Barbut F, Menuet D, Verachten M, Girou E. Comparison of the efficacy of a hydrogen peroxide dry-mist disinfection system and sodium hypochlorite solution for eradication of Clostridium difficile spores. *Infection Control Hosp Epidemiol*. (2009) 30:507–14. doi: 10.1086/597232
19. Holmdahl T, Lanbeck P, Wullt M, Walder M. A head-to-head comparison of hydrogen peroxide vapor and aerosol room decontamination systems. *Infection Control Hosp Epidemiol*. (2011) 32:831–6. doi: 10.1086/661104
20. Fu T, Gent P, Kumar V. Efficacy, efficiency and safety aspects of hydrogen peroxide vapour and aerosolized hydrogen peroxide room disinfection systems. *J Hosp Infect*. (2012) 80:199–205. doi: 10.1016/j.jhin.2011.11.019
21. Haas JP, Menz J, Dusza S, Montecalvo MA. Implementation and impact of ultraviolet environmental disinfection in an acute care setting. *Am J Infect Control*. (2014) 42:586–90. doi: 10.1016/j.ajic.2013.12.013
22. Memarzadeh F, Olmsted RN, Bartley JM. Applications of ultraviolet germicidal irradiation disinfection in health care facilities: effective adjunct, but not stand-alone technology. *Am J Infect Control*. (2010) 38:S13–S24. doi: 10.1016/j.ajic.2010.04.208
23. Lu H. Polarization separation by an adaptive filter. *IEEE Trans Aerosp Electron Syst*. (1973) 6:954–6. doi: 10.1109/TAES.1973.309675
24. Urciuoli A, Buono A, Nunziata F, Migliaccio M. Analysis of local-and non-local filters for multi-polarization SAR coastline extraction applications. In: *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*. Florence: IEEE (2019). p. 28–33.
25. Lee JS, Grunes MR, De Grandi G. Polarimetric SAR speckle filtering and its implication for classification. *IEEE Trans Geosci Remote Sens*. (1999) 37:2363–73. doi: 10.1109/36.789635
26. Qu T, Wang X, Zhao F. Effect of many air cooling fans in operation on dynamic response of bridge truss. In: *2011 International Conference on Consumer Electronics, Communications and Networks (CECNet)*. Xianning: IEEE (2011). p. 1194–7.
27. Calisti M, Carbonara G, Laschi C. A rotating polarizing filter approach for image enhancement. In: *OCEANS 2017-Aberdeen*. Aberdeen: IEEE (2017). p. 1–4.
28. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
29. Carl R, Lawless P, Decker J. Characterization and mitigation of cooling fan installation penalties in 1U enterprise class servers. In: *2014 Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM)*. San Jose, CA: IEEE (2014). p. 213–8.
30. Huang Z, Cao Y, Wang T. Transfer learning with efficient convolutional neural networks for fruit recognition. In: *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. Chengdu: IEEE (2019). p. 358–62.
31. Pande A, Munot M, Sreeemathy R, Bakare R. An efficient approach to fruit classification and grading using deep convolutional neural network. In: *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*. Bombay: IEEE (2019). p. 1–7.
32. Zhu D, Wang M, Zou Q, Shen D, Luo J. Research on fruit category classification based on convolution neural network and data augmentation. *In: 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*. Xiamen: IEEE (2019). p. 46–50.
33. Kausar A, Sharif M, Park J, Shin DR. Pure-cnn: a framework for fruit images classification. In: *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. Las Vegas, NV: IEEE (2018). p. 404–8.
34. Das S, Ganguly S, Ghosh S, Sarker R, Sengupta D. A bluetooth based sophisticated home automation system using smartphone. In: *2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI)*. Kolkata: IEEE (2016). p. 236–40.
35. Kumar S, Lee SR. Android based smart home system with control via Bluetooth and internet connectivity. In: *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*. Jeju: IEEE (2014). p. 1–2.
36. Singh A, Gupta T, Korde M. Bluetooth controlled spy robot. In: *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*. Indore: IEEE (2017). p. 1–4.
37. Paala NMA, Garcia NMM, Supetran RA, Fontamillas MELB. Android controlled lawn mower using Bluetooth and WIFI connection. In: *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*. Singapore: IEEE (2019). p. 702–6.
38. Shinde A, Kanade S, Jugale N, Gurav A, Vatti RA, Patwardhan M. Smart home automation system using IR, bluetooth, GSM and android. In: *2017 Fourth International Conference on Image Information Processing (ICIIP)*. Shimla: IEEE (2017). p. 1–6.

39. Tripti NF, Farhad A, Iqbal W, Zaman HU. SaveMe: a crime deterrent personal safety android app with a bluetooth connected hardware switch. In: *2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC)*. Shah Alam: IEEE (2018). p. 23–6.
40. Agarwal S, Agarwal N. Interfacing of robot with android app for to and fro communication. In: *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls With Their Impact on Humanity (CIPECH)*. Ghaziabad: IEEE (2016). p. 222–6.
41. Fruit-360 dataset. (2021). Available online at: https://www.kaggle.com/moltean/fruits.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kumar, Bansal, Alroobaea, Baqasah and Hedabou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fpubh.2021.831940

Development and Validation of a Nomogram to Predict Distant Metastasis in Elderly Patients With Renal Cell Carcinoma

Jinkui Wang[1], *Chenghao Zhanghuang*[1,2], *Xiaojun Tan*[1,3], *Tao Mi*[1], *Jiayan Liu*[1], *Liming Jin*[1], *Mujie Li*[1], *Zhaoxia Zhang*[1] *and Dawei He*[1]*

[1] *Department of Urology, Chongqing Key Laboratory of Children Urogenital Development and Tissue Engineering, Chongqing Key Laboratory of Pediatrics, Ministry of Education Key Laboratory of Child Development and Disorders, China International Science and Technology Cooperation Base of Child development and Critical Disorders, National Clinical Research Center for Child Health and Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China,* [2] *Department of Urology, Yunnan Key Laboratory of Children's Major Disease Research, Kunming Children's Hospital (Children's Hospital Affiliated to Kunming Medical University), Kunming, China,* [3] *Department of Urology, Nanchong Central Hospital, The Second Clinical Medical College, North Sichuan Medical University, Nanchong, China*

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Abdul Rehman Javed, Air University, Pakistan

***Correspondence:**
Dawei He
hedawei@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 December 2021*
Accepted: *24 December 2021*
Published: *28 January 2022*

Citation:
Wang J, Zhanghuang C, Tan X, Mi T, Liu J, Jin L, Li M, Zhang Z and He D (2022) Development and Validation of a Nomogram to Predict Distant Metastasis in Elderly Patients With Renal Cell Carcinoma. Front. Public Health 9:831940. doi: 10.3389/fpubh.2021.831940

Background: Renal cell carcinoma (RCC) is the most common renal malignant tumor in elderly patients. The prognosis of renal cell carcinoma with distant metastasis is poor. We aim to construct a nomogram to predict the risk of distant metastasis in elderly patients with RCC to help doctors and patients with early intervention and improve the survival rate.

Methods: The clinicopathological information of patients was downloaded from SEER to identify all elderly patients with RCC over 65 years old from 2010 to 2018. Univariate and multivariate logistic regression analyzed the training cohort's independent risk factors for distant metastasis. A nomogram was established to predict the distant metastasis of elderly patients with RCC based on these risk factors. We used the consistency index (C-index), calibration curve, and area under the receiver operating curve (AUC) to evaluate the accuracy and discrimination of the prediction model. Decision curve analysis (DCA) was used to assess the clinical application value of the model.

Results: A total of 36,365 elderly patients with RCC were included in the study. They were randomly divided into the training cohort ($N = 25,321$) and the validation cohort ($N = 11,044$). In the training cohort, univariate and multivariate logistic regression analysis suggested that race, tumor histological type, histological grade, T stage, N stage, tumor size, surgery, radiotherapy, and chemotherapy were independent risk factors for distant metastasis elderly patients with RCC. A nomogram was constructed to predict the risk of distant metastasis in elderly patients with RCC. The training and validation cohort's C-indexes are 0.949 and 0.954, respectively, indicating that the nomogram has excellent accuracy. AUC of the training and validation cohorts indicated excellent predictive ability. DCA suggested that the nomogram had a better clinical application value than the traditional TN staging.

Conclusion: This study constructed a new nomogram to predict the risk of distant metastasis in elderly patients with RCC. The nomogram has excellent accuracy and reliability, which can help doctors and patients actively monitor and follow up patients to prevent distant metastasis of tumors.

Keywords: nomogram, renal cell carcinoma, risk, distant metastasis, SEER

BACKGROUND

Renal cell carcinoma (RCC) is the most common renal malignant tumor in adults, accounting for about 3% of all human tumors (1), with more than 400,000 newly diagnosed patients each year (2). The proportion of elderly patients over 65 years old in RCC is more than 70% (3). Moreover, due to the increase of population aging and the extension of life expectancy caused by improved medical levels, the proportion is still rising (1).

According to the tumor metastasis, RCC is divided into metastatic RCC (mRCC) and non-metastatic RCC (nmRCC). Tumor metastasis is the critical factor to determine the prognosis of RCC patients. The prognosis of patients with mRCC is poor, and the median survival time is only 10.2 months (4), while the prognosis of nmRCC is good, and it is even considered to be cured entirely (5). Studies have shown that 18–30% of RCC patients have distant metastasis at the first diagnosis (6). In comparison, 33% of patients have recurrence and metastasis after surgical excision of the tumor, and it is unclear which patients are prone to metastasis (7).

The main sites of RCC metastasis were lung and bone, accounting for 43.6 and 27.6% of mRCC, respectively. Brain and liver metastasis accounted for 4.4%, respectively (8). Data processing technology has been continuously improved by continuously updating databases such as surveillance, epidemiology, and final results (SEER) of the National Cancer Institute of the United States. The nomograms prediction model represented by UISS (9), SSIGN (10), Leibovich (11) was born and provided treatment suggestions for clinicians. In terms of the RCC transfer prediction model, we found that the prediction models for each leading transfer site of RCC were reported (12–14), but the C-index of the model was 0.714–0.803, and the area under the curve (AUC) was 0.767–0.780, indicating that the relative accuracy was insufficient. And the distant metastasis prediction model for the overall large sample of the elderly has not been reported.

At present, artificial intelligence has been widely used in human health. Dhanamjayulu et al. (15) used real-time image processing and machine learning to identify malnutrition, predict BMI from facial images, and identify unhealthy people quickly. Gadekallu et al. (16, 17) used the deep learning model for the early detection of diabetic retinopathy. And the nomogram is the most commonly used prediction model. To improve the prognosis of elderly patients with RCC, we developed a risk prediction model for distant metastasis of elderly RCC. We validated its accuracy to provide treatment guidance for clinical work.

METHOD

Data Source and Data Extraction

The patient information was extracted from the Surveillance, Epidemiology, and End Results (SEER) project of the National Cancer Institute in the United States to include all elderly patients with RCC from 2010 to 2018. SEER database is the National Cancer Database of the United States, including 18 cancer registration centers, covering about 28% of the national population (18). Demographic information, clinicopathological information, and follow-up data of cancer patients can be downloaded from the SEER database. Access to http://seer.cancer.gov/ provides all the data for this study. Since the patient information in the SEER database is public and anonymous, our research does not require ethical approval and patients' informed consent. Our research method conforms to the research criteria of SEER data release.

We collected patient information, including age, sex, race, year of diagnosis, marital status, tumor histological type, histological grade, tumor side, tumor size, T stage, N stage, surgical method, radiotherapy, and chemotherapy. Inclusion criteria: (1) pathological diagnosis of RCC (ICD-O-3 code 8260, 8310, 8312, 8317); (2) Aged 65 or above; (3) The years of diagnosis were 2010–2018. Exclusion criteria: (1) the race of the patient is unknown; (2) bilateral or unilateral renal tumor; (3) TN staging is unknown; (4) incomplete follow-up information; (5) tumor size is unknown; (6) unknown surgical method; (7) survival time <1 month. The patient screening flow chart is shown in **Figure 1**.

Patients' race was divided into three categories: white, black, and others (American Indian/AK Native, Asian/Pacific Islander). The patients' years of diagnosis were divided into 2010–2014, 2015–2018. Marital status is divided into married and unmarried (including single, divorced, and widowed). Histological tumors include renal clear cell carcinoma, renal papillary cell carcinoma, chromophobe cell carcinoma, and some unclassified RCC. The histological grading of tumors is classified as grades I–IV, which are well-differentiated, moderately differentiated, poorly differentiated, and undifferentiated. Surgical procedures were classified as local excision of tumor (SEER codes 10–27), partial nephrectomy (SEER codes 30), and radical nephrectomy (SEER codes 40–80).

Construction and Validation of the Nomogram

All patients enrolled were randomly assigned to either the training cohort (70%) or the validation cohort (30%). In the training cohort, univariate, and multivariate logistic regression models were used to analyze the independent risk factors for

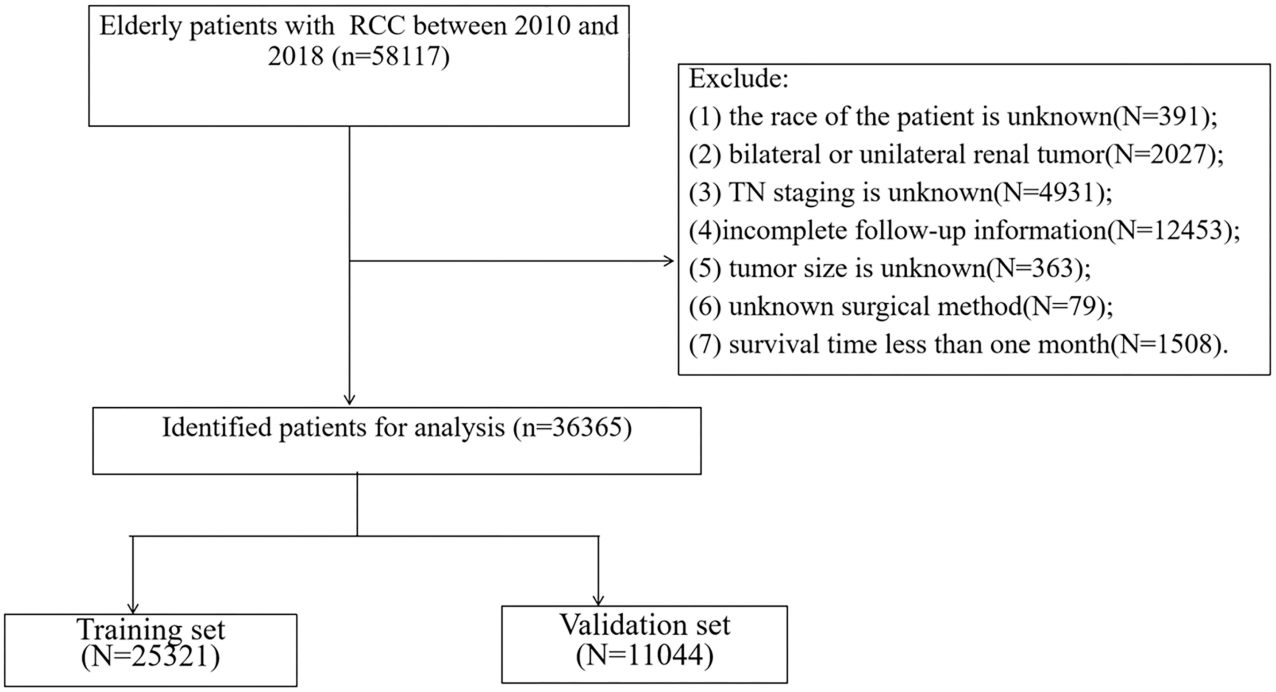


FIGURE 1 | The flowchart of including and dividing patients.

patient metastasis, and hazard ratio (HR) and 95% confidence interval (CI) were recorded simultaneously. The independent risk factors screened were used to establish a nomogram to predict the risk of distant metastasis in elderly patients with RCC. We validated the accuracy of the prediction model of the training cohort and the validation cohort by the calibration curve. We used the consistency index (C-index) to demonstrate the discrimination ability of the model. Meanwhile, We used the area under the receiver operating curve (AUC) to validate the model's accuracy.

Clinical Utility

Decision curve analysis (DCA) was used to validate the clinical value of the model. DCA is a new algorithm to evaluate the net benefit value of the model under different thresholds (19). We also used DCA to compare the ability of the nomogram and TN staging to predict patients' risk of distant metastasis. In addition, we used a risk stratification system to divide all patients into high-risk and low-risk groups based on their nomogram scores. Log-rank test and Kaplan-Meier (K-M) curve were used to compare the survival and prognosis of patients in different risk groups.

Statistical Analysis

Frequency description (%) was used for counting data, and the chi-square test or non-parametric *U*-test was used to compare groups. Measurement data (age, tumor size) were described using mean and standard deviation, and non-parametric *U*-tests were used to compare groups. Univariate and multivariate logistic regression models analyzed the risk factors of distant metastasis. The K-M curve and log-rank test compared the survival differences between groups. All statistical analyses were performed using R software 4.1.0 (http://www.Rproject.org) and SPSS 26.0 (IBM, Chicago, IL, USA). $P < 0.05$ was considered statistically significant.

RESULTS

Clinical Features

we enrolled 36,365 elderly patients with RCC based on inclusion and exclusion criteria. The mean age of the patients was 73.4 ± 6.58 years, and 29,891 (82.2%) were white, 22,760 (62.6%) were male, and 21,817 (60.0%) were married. There were 2,772 patients (7.62%), 12,432 patients (34.2%), 7,630 patients (21.0%), and 1,778 patients (4.89%) of histological tumor grade I-IV, and 20,279 patients (55.8%) of tumor histological type RCC. Papillary cell carcinoma was 5,207 (14.3%), and chromophobe cell carcinoma was 1,855 (5.10%). The mean tumor size was 49.7 ± 34.5 mm. T stage included 17,089 (47.0%) T1a, 8,320 (22.9%) T1b, 3,475 (9.56%) T2, 7,287 (20.0%) T3, and 194 (0.53%) T4. Thirty-four thousand seven hundred and fifteen (95.5%) patients had stage N0. There were 7,254 (19.9%) patients without surgery, 3,083 (8.48%) patients with local tumor excision, 3,083 (8.48%) patients with partial nephrectomy, and 16,266 (44.7%) patients with radical nephrectomy. There were 2,269 (6.24%) patients who received chemotherapy and 1,105 (3.04%) patients who received radiotherapy. The clinicopathological information of all patients is shown in **Table 1**. There was no significant difference between the training cohort and the validation cohort.

TABLE 1 | Clinicopathological information in elderly patients with renal cell carcinoma.

	ALL *N* = 36,365	Training cohort *N* = 25,321	Validation cohort *N* = 11,044	*p*
Age	73.4 (6.58)	73.4 (6.58)	73.3 (6.58)	0.227
Race				0.980
White	29,891 (82.2%)	20,807 (82.2%)	9,084 (82.3%)	
Black	4,279 (11.8%)	2,982 (11.8%)	1,297 (11.7%)	
Other	2,195 (6.04%)	1,532 (6.05%)	663 (6.00%)	
Sex				0.677
Male	22,760 (62.6%)	15,866 (62.7%)	6,894 (62.4%)	
Female	13,605 (37.4%)	9,455 (37.3%)	4,150 (37.6%)	
Year of diagnosis				0.054
2010–2014	18,953 (52.1%)	13,112 (51.8%)	5,841 (52.9%)	
2015–2018	17,412 (47.9%)	12,209 (48.2%)	5,203 (47.1%)	
Marital				0.355
No	14,548 (40.0%)	10,170 (40.2%)	4,378 (39.6%)	
Married	21,817 (60.0%)	15,151 (59.8%)	6,666 (60.4%)	
Histologic type				0.337
Clear cell	20,279 (55.8%)	14,113 (55.7%)	6,166 (55.8%)	
Papillary	5,207 (14.3%)	3,621 (14.3%)	1,586 (14.4%)	
Chromophobe	1,855 (5.10%)	1,261 (4.98%)	594 (5.38%)	
Not classified	9,024 (24.8%)	6,326 (25.0%)	2,698 (24.4%)	
Grade				0.537
I	2,772 (7.62%)	1,934 (7.64%)	838 (7.59%)	
II	12,432 (34.2%)	8,605 (34.0%)	3,827 (34.7%)	
III	7,630 (21.0%)	5,346 (21.1%)	2,284 (20.7%)	
IV	1,778 (4.89%)	1,219 (4.81%)	559 (5.06%)	
Unknown	11,753 (32.3%)	8,217 (32.5%)	3,536 (32.0%)	
Laterality				0.970
Left	18,015 (49.5%)	12,546 (49.5%)	5,469 (49.5%)	
Right	18,350 (50.5%)	12,775 (50.5%)	5,575 (50.5%)	
T				0.228
T1a	17,089 (47.0%)	11,907 (47.0%)	5,182 (46.9%)	
T1b	8,320 (22.9%)	5,844 (23.1%)	2,476 (22.4%)	
T2	3,475 (9.56%)	2,435 (9.62%)	1,040 (9.42%)	
T3	7,287 (20.0%)	4,998 (19.7%)	2,289 (20.7%)	
T4	194 (0.53%)	137 (0.54%)	57 (0.52%)	
N				0.561
N0	34,715 (95.5%)	24,161 (95.4%)	10,554 (95.6%)	
N1	1,650 (4.54%)	1,160 (4.58%)	490 (4.44%)	
Surgery				0.346
No	7,254 (19.9%)	5,092 (20.1%)	2,162 (19.6%)	
Local tumor excision	3,083 (8.48%)	2,132 (8.42%)	951 (8.61%)	
Partial nephrectomy	9,762 (26.8%)	6,835 (27.0%)	2,927 (26.5%)	
Radical nephrectomy	16,266 (44.7%)	11,262 (44.5%)	5,004 (45.3%)	
Chemotherapy				0.439
No/Unknown	34,096 (93.8%)	23,758 (93.8%)	10,338 (93.6%)	
Yes	2,269 (6.24%)	1,563 (6.17%)	706 (6.39%)	
Radiation				0.085
No/Unknown	35,260 (97.0%)	24,578 (97.1%)	10,682 (96.7%)	
Yes	1,105 (3.04%)	743 (2.93%)	362 (3.28%)	
Tumor size	49.7 (34.5)	49.6 (34.2)	50.0 (35.3)	0.426

(Continued)

TABLE 1 | Continued

	ALL *N* = 36,365	Training cohort *N* = 25,321	Validation cohort *N* = 11,044	*p*
Bone metastasis				0.063
No/Unknown	34,965 (96.2%)	24,378 (96.3%)	10,587 (95.9%)	
Yes	1,400 (3.85%)	943 (3.72%)	457 (4.14%)	
Brain metastasis				0.615
No/Unknown	36,025 (99.1%)	25,089 (99.1%)	10,936 (99.0%)	
Yes	340 (0.93%)	232 (0.92%)	108 (0.98%)	
Liver metastasis				0.666
No/Unknown	35,768 (98.4%)	24,900 (98.3%)	10,868 (98.4%)	
Yes	597 (1.64%)	421 (1.66%)	176 (1.59%)	
Lung metastasis				0.060
No/unknown	34,293 (94.3%)	23,917 (94.5%)	10,376 (94.0%)	
Yes	2,072 (5.70%)	1,404 (5.54%)	668 (6.05%)	

Univariate and Multivariate Logistic Regression Analysis

In the training cohort, using the univariate logistic regression model to screen a risk factor for distant metastasis, the results showed that age, gender, race, year of diagnosis, marriage, tumor histologic type, histologic grade, T stage, N stage, tumor size, surgery, radiation therapy, chemotherapy is a risk factor for RCC distant metastases. The multivariate logistic regression model analyzed these variables. The results showed that race, tumor histological type, histological grade, T stage, N stage, tumor size, surgery, radiotherapy, and chemotherapy were independent risk factors for distant metastasis in elderly patients with RCC (**Table 2**). We can incorporate these risk factors into a nomogram to predict the risk of distant metastasis.

Construction and Validation of the Nomogram

Based on independent risk factors derived from univariate and multivariate screening, we constructed a new nomogram to predict the risk of distant metastasis in elderly patients with RCC (**Figure 2**). Tumor size was the most significant risk factor for distant metastasis, followed by radiotherapy, surgery, T stage, chemotherapy, histological grade, histological type, N stage, and race. Subsequently, calibration curves were used to validate the accuracy of the nomogram in the training cohort and the validation cohort, respectively. The results showed that the model's predicted values were consistent with the actual observed values (**Figure 3**). In the training and validation cohorts, the C-index was 0.949 (95% CI, 0.945–0.953) and 0.954 (95% CI, 0.947–0.960), respectively, indicating excellent accuracy. The AUC of the training cohort and the validation cohort also suggested excellent predictive power, with 0.95 (95%CI, 0.945–0.954) and 0.954 (95%CI, 0.947–0.96), respectively (**Figure 4**).

Clinical Application of Nomogram

DCA showed good clinical utility in training and validation cohorts (**Figure 5**). In addition, the nomogram has an obvious predictive advantage over TN staging. Patients in the training cohort and validation cohort were divided into the high-risk group (total score $>$ 99.7) and low-risk group ($\leq$99.7 overall). The K-M curve indicated that the 1-, 3-, and 5-year survival rates in the high-risk group were 0.627 (95% CI, 0.613–0.642), 0.406 (95% CI, 0.389–0.423), and 0.300 (95% CI, 0.282–0.320), respectively. The 1-, 3- and 5-year survival rates of patients in the low-risk group were 0.967 (95% CI, 0.964–0.969), 0.922 (95% CI, 0.919–0.926), and 0.880 (95% CI, 0.876–0.885), respectively (**Figure 6**). The results showed that patients in the high-risk group had a higher risk of distant metastasis and a worse prognosis. We analyzed surgical procedures in the high-risk and low-risk groups. In the low-risk group, patients with partial nephrectomy had the highest survival rate, followed by local tumor excision and radical nephrectomy. Patients without surgery had the worst survival rate (**Figure 7A**). In the high-risk group, where most patients did not undergo surgery or underwent radical nephrectomy, survival was significantly higher among patients undergoing surgery than among patients who did not undergo surgery (**Figure 7B**).

DISCUSSION

RCC is a high incidence tumor, with an estimated 330,000 patients worldwide every year (20), and RCC is still a severe problem that needs to be addressed in the United States (21). Non-metastatic RCC has a good prognosis after surgical treatment, including partial nephrectomy or radical nephrectomy (22). However, 30% of patients eventually develop metastatic RCC after surgical treatment, which is incurable (23). Currently, treatment for patients with metastatic RCC is minimal, including the use of cancer-targeted surgery and immune checkpoint inhibitor drugs to treat metastatic RCC. Still, the effect is inferior (24, 25). At present, CT is mainly used to detect the metastasis of RCC. Still, CT also has its inherent limitations, and its prediction of the risk of cancer metastasis is minimal (26). Therefore, it is necessary to accurately predict the risk of metastasis in patients, strengthening doctors' active monitoring of patients and preventing distant metastasis early.

TABLE 2 | Univariate and multivariate analyses of CSS in training set.

	Univariate			Multivariate		
	HR	**95% CI**	***P***	**HR**	**95% CI**	***P***
Age	1.02	1.01–1.02	<0.001	0.992	0.983–1	0.062
Race						
White	Reference					
Black	0.73	0.63–0.85	0	0.869	0.706–1.07	0.187
Other	1.14	0.97–1.35	0.11	1.358	1.072–1.72	0.011
Sex						
Male	Reference					
Female	0.87	0.8–0.95	<0.001			
Year of diagnosis						
2010–2014						
2015–2018	1.12	1.03–1.22	0.01			
Marital	Reference					
No						
Married	0.92	0.84–1	0.04			
Histologic type						
Clear cell	Reference					
Papillary	0.36	0.29–0.43	0	0.45	0.343–0.591	0
Chromophobe	0.21	0.14–0.32	0	0.262	0.154–0.445	0
Not classified	2.12	1.94–2.32	0	0.789	0.668–0.933	0.006
Grade						
I						
II	2.08	1.41–3.06	0	1.915	1.16–3.161	0.011
III	5.51	3.77–8.06	0	3.17	1.926–5.219	0
IV	20.69	14.02–30.55	0	5.565	3.315–9.342	0
Unknown	13.97	9.64–20.24	0	4.066	2.498–6.616	0
Laterality						
Left						
Right	0.95	0.87–1.03	0.22			
T						
T1a	Reference					
T1b	3.48	2.99–4.05	0	2.614	2.132–3.206	0
T2	11.14	9.55–12.98	0	4.639	3.616–5.953	0
T3	11.79	10.29–13.5	0	6.758	5.378–8.492	0
T4	74.58	51.57–107.85	0	14.282	8.417–24.236	0
N						
N0						
N1	21.81	19.18–24.81	0	3.858	3.219–4.622	0
Surgery						
No	Reference					
Local tumor excision	0.02	0.01–0.03	0	0.112	0.067–0.185	0
Partial nephrectomy	0.01	0.01–0.02	0	0.052	0.036–0.077	0
Radical nephrectomy	0.19	0.17–0.21	0	0.138	0.112–0.17	0
Chemotherapy:						
No/Unknown						
Yes	52.27	46.02–59.37	0	12.437	10.58–14.62	0
Radiation						
No/unknown						
Yes	83.54	67.18–103.87	0	28.853	22.196–37.506	0
Tumor size	1.03	1.02–1.03	0	1.006	1.004–1.008	0

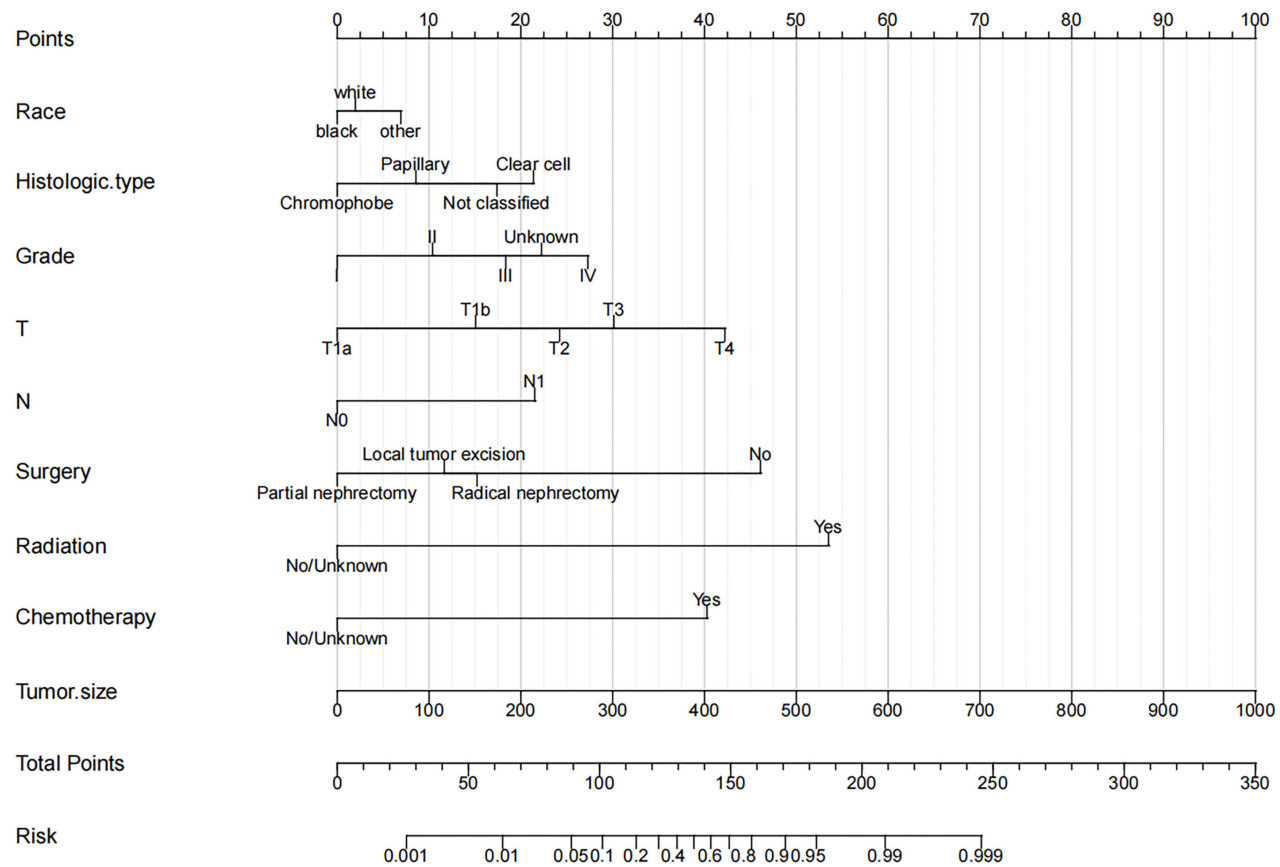


FIGURE 2 | Nomogram for distant metastasis of elderly patients with RCC.

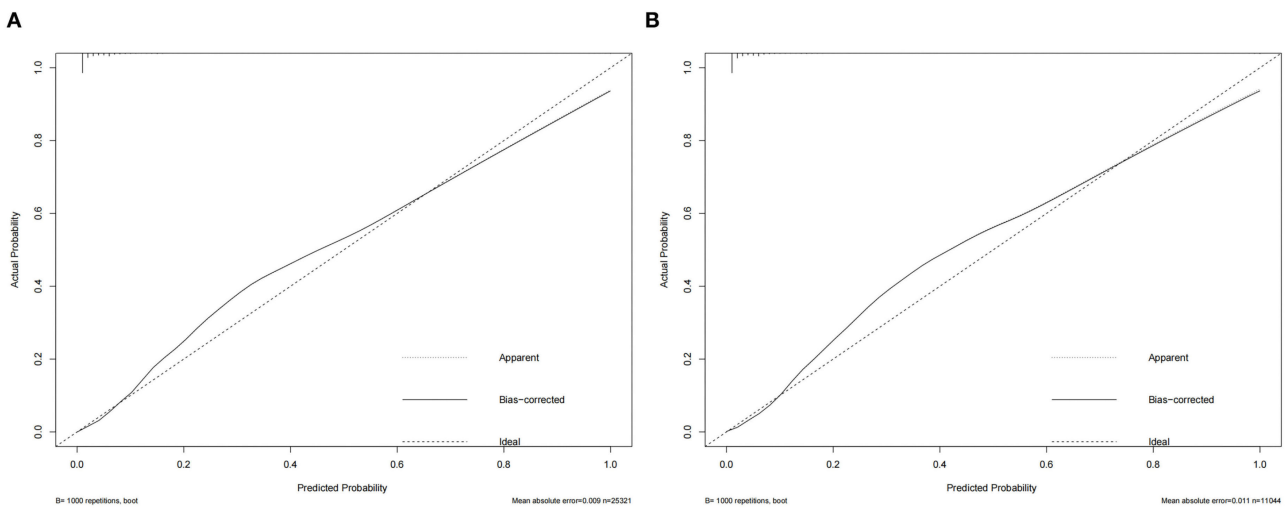


FIGURE 3 | The nomogram's calibration curve in the training cohort **(A)** and the validation cohort **(B)**.

At present, it is very limited in predicting distant metastasis of RCC. Hutterer et al. (27) collected renal cell carcinoma patients in 12 medical centers and established rosettes to predict distant lymph node metastasis, with an accuracy of 78.4%. Capitanio et al. (28) also established a prediction model to predict lymph node metastasis in RCC patients, with an accuracy of 86.9%.

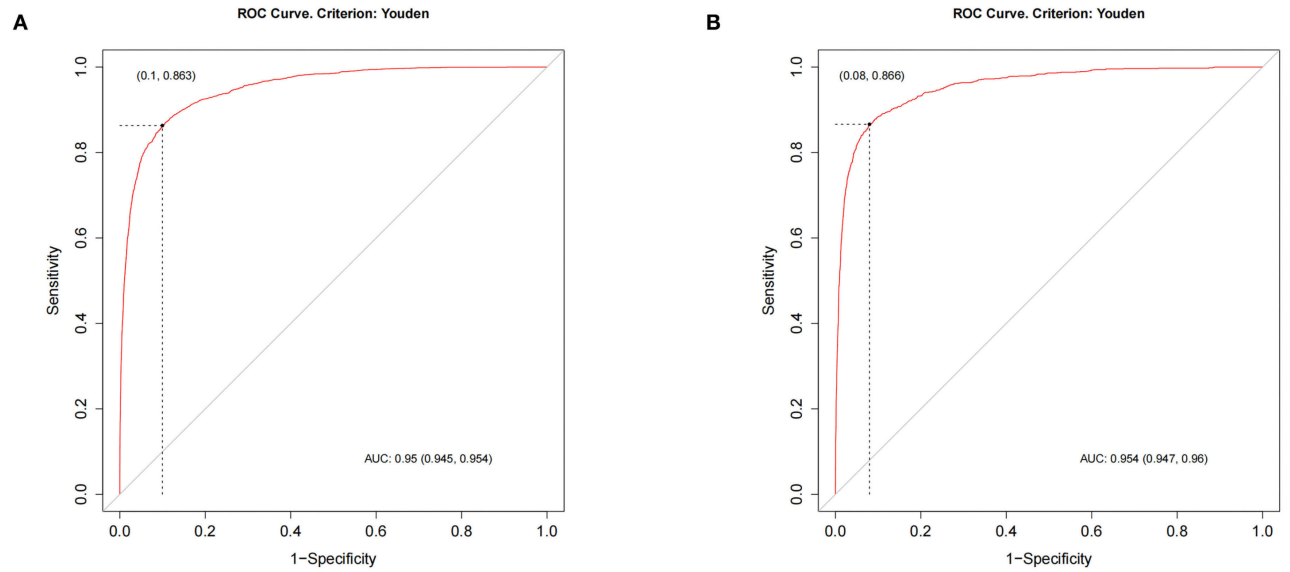


FIGURE 4 | The ROC of the nomogram of training cohort **(A)** and validation cohort **(B)**.

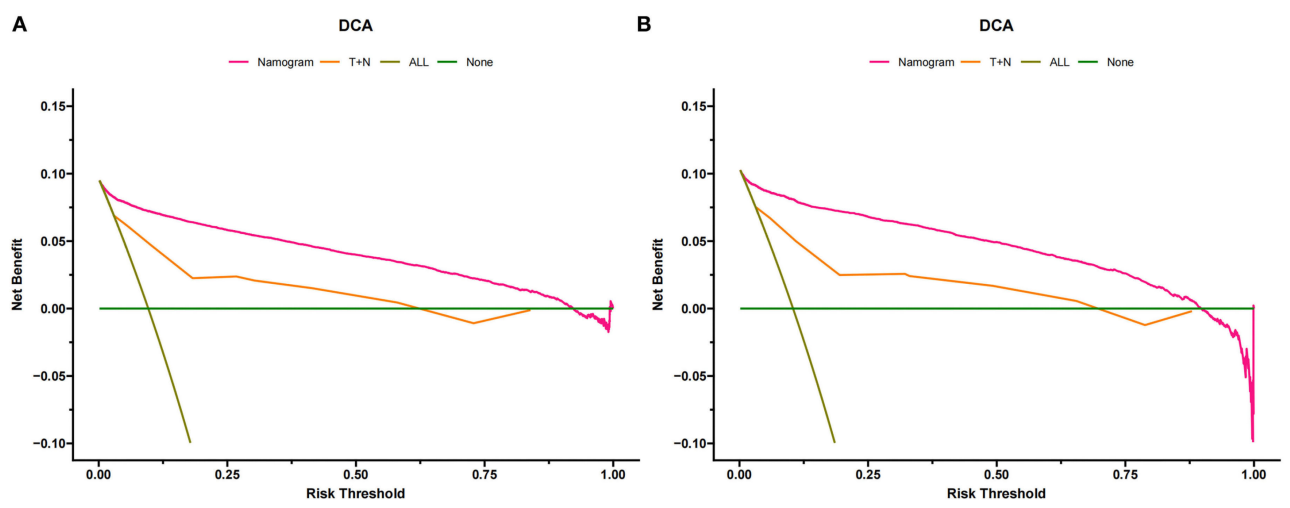


FIGURE 5 | (A) Decision curves of the nomogram predicting distant metastasis in the training cohort and **(B)** the validation cohort. The x-axis is the risk threshold, and the y-axis is the net benefit. The purple line indicates no distant metastasis of the patient, and the blue line shows that all patients have metastasis. When the threshold probability is between 0 and 100%, the net benefit of the model is the largest.

Marconi et al. (29) established a prediction model to predict the survival rate of patients with distant metastasis. Bai et al. (30) used MRI radiomic-based nomogram to predict distant metastasis of patients. Similarly, Zhao et al. (31) used CT radiomics to predict distant metastasis of patients. In addition, Li et al. (32) used patients in the SEER database to predict distant metastasis in RCC patients, but the C index was only 0.863. As far as we know, the current prediction model for distant metastases in RCC patients using nomograms has defects of a small number of patients and low prediction accuracy (**Supplementary Table 1**). Therefore, we aim to construct a highly accurate predictive model to predict distant metastases in elderly patients with RCC. The C-index of our prediction model is 0.949, which is much higher than the previous prediction models.

This study explored the risk factors for distant metastasis of elderly patients with RCC. We found that tumor size is the leading risk factor for distant metastasis of RCC, and the larger the tumor, the higher the risk of distant metastasis. Our results are similar to previous studies. Hutterer et al. (33) previously established a nomogram to predict distant metastasis of RCC and found that tumor size was a significant risk factor. Zastrow

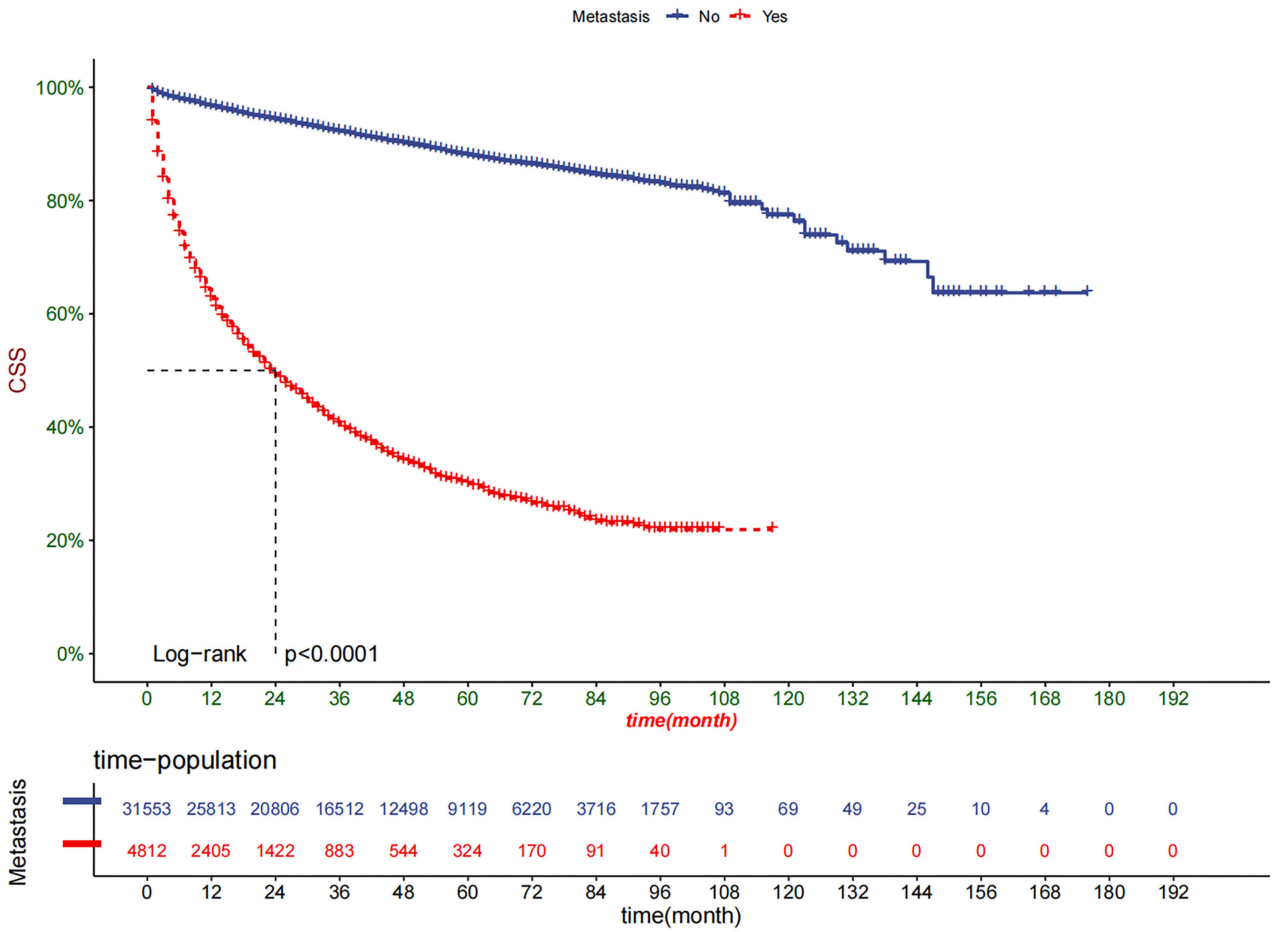


FIGURE 6 | Kaplan–Meier curves of CSS for patients in the low-risk and high-risk groups.

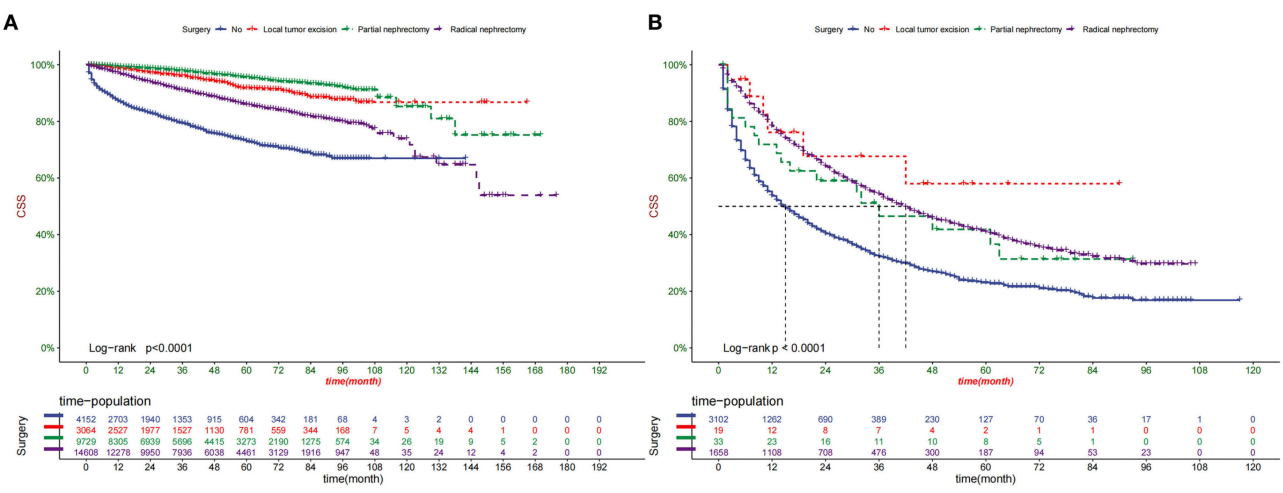


FIGURE 7 | (A) Kaplan–Meier curves of CSS for patients with different surgery in the low-risk group and **(B)** the high-risk group.

et al. (34) also found that tumor size was a risk factor for distant metastasis of renal cells. The risk of distant metastasis was significantly increased when tumor size was more significant than 3 cm. Li et al. constructed a nomogram to predict distant metastasis of RCC and found that tumors larger than 3 cm would increase the risk of metastasis (32).

Our study also found that tumor histological type and grade are important risk factors for distant metastasis. Patients with clear cell carcinoma of the kidney had the highest risk of distant metastasis, followed by papillary cell carcinoma of the kidney, and chromophobe cell carcinoma had the lowest risk of metastasis. Our results were similar to those of Zastrow et al. (34), who found that the risk of distant metastasis of renal clear cell carcinoma was the highest. In addition, we found that the degree of differentiation of RCC was associated with distant metastasis. The worse the differentiation, the higher the risk of metastasis. Previous studies have also found this phenomenon (34, 35). Because less differentiated tumors represent more aggressive biological behavior, which means they are more likely to metastasize far away.

In addition, the T stage and N stage are also crucial factors for tumor metastasis. Our study found that the higher the T stage, the greater the risk of tumor metastasis. This is understandable because the tumor invades the blood vessels and causes cancer cells to enter the bloodstream and metastasize. Lymph node metastasis is also a risk factor for distant metastasis of tumors. Previous studies have proved that regional lymph node involvement can lead to distant metastasis of tumors (6). They confirmed that regional lymph node involvement could increase the risk of distant metastasis by 50%.

In addition, surgery is also a key factor for distant metastasis of tumors. Patients without surgery are more likely to have distant metastasis than those with surgery because surgery can effectively remove or destroy cancer, thus reducing the chance of tumor metastasis. We divided patients into the high-risk metastatic and low-risk groups, and in the high-risk group, patients who underwent surgery had significantly improved survival outcomes. Patients with partial nephrectomy had the highest survival rate in the low-risk group. Because there are many complications in elderly patients, partial nephrectomy can remove the tumor with enough nephron remaining, resulting in a higher survival rate (36). Therefore, in the absence of contraindications, surgical treatment is still recommended to achieve a better outcome for patients at high risk of metastasis. For low-risk patients, partial nephrectomy is the best option.

This study explored the risk factors for distant metastasis in elderly patients with RCC. We used these risk factors to construct a graph to predict distant metastasis. After internal validation, the C-index could reach 0.95, and both the calibration curve and DCA showed the excellent accuracy of the prediction model. This prediction model can provide a practical, theoretical basis for patients' clinical decision-making and postoperative follow-up. To improve the survival rate and quality of life of elderly patients with RCC, we can better monitor the population at high risk of distant metastasis.

However, there are still some limitations to this study. First of all, this study is a retrospective study, so it is challenging to avoid selection bias. Therefore, prospective clinical trials are necessary to test the accuracy of the prediction model. Secondly, the SEER database lacks many related risk factors, such as comorbidity information, smoking, hypertension, BMI, etc. However, we included essential determinants such as tumor stage, histological type, and size so that our prediction model could achieve remarkably high accuracy. Finally, our prediction model is only validated internally, and further external validation is necessary to validate the accuracy and reliability of the model. Next, we plan to conduct a multi-center prospective clinical study to verify the accuracy of this prediction model.

CONCLUSION

This study explored the risk factors for distant metastasis in elderly patients with RCC and found that race, tumor histological type, histological grade, T stage, N stage, tumor size, surgery, radiotherapy, and chemotherapy were independent risk factors for distant metastasis. We constructed a new nomogram to predict the risk of distant metastasis in elderly patients with RCC. With good accuracy and reliability, this nomogram can help doctors and patients to carry out active monitoring and follow-up of patients to prevent distant metastasis of tumors.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov/.

ETHICS STATEMENT

The data of this study is obtained from the SEER database. The patients' data is public and anonymous, so this study does not require ethical approval and informed consent.

AUTHOR CONTRIBUTIONS

JW and CZ designed the study. CZ, JW, JL, LJ, and XT collected and analyzed the data. JW drafted the initial manuscript. CZ, TM, JL, ZZ, and DH revised the article critically. CZ, JL, DH, LJ, and XT reviewed and edited the paper. All authors approved the final manuscript.

FUNDING

This work was supported by Special Key Project of Chongqing Technology Innovation and Application Development (No. Cstc2019jscx-tjsbX0003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.831940/full#supplementary-material

REFERENCES

1. González León T, Morera Pérez M. Renal cancer in the elderly. *Curr Urol Rep.* (2016) 17:6. doi: 10.1007/s11934-015-0562-2
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [published correction appears in CA Cancer J Clin. 2020 Jul;70(4):313]. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
3. Quivy A, Daste A, Harbaoui A, Duc S, Bernhard JC, Gross-Goupil M, et al. Optimal management of renal cell carcinoma in the elderly: a review. *Clin Interv Aging.* (2013) 8:433–42. doi: 10.2147/CIA.S30765
4. Schwaab T, Schwarzer A, Wolf B, Crocenzi TS, Seigne JD, Crosby NA, et al. Clinical and immunologic effects of intranodal autologous tumor lysate-dendritic cell vaccine with Aldesleukin (Interleukin 2) and IFN-{alpha}2a therapy in metastatic renal cell carcinoma patients. *Clin Cancer Res.* (2009) 15:4986–92. doi: 10.1158/1078-0432.CCR-08-3240
5. Kutikov A, Egleston BL, Wong YN, Uzzo RG. Evaluating overall survival and competing risks of death in patients with localized renal cell carcinoma using a comprehensive nomogram. *J Clin Oncol.* (2010) 28:311–7. doi: 10.1200/JCO.2009.22.4816
6. Tadayoni A, Paschall AK, Malayeri AA. Assessing lymph node status in patients with kidney cancer. *Transl Androl Urol.* (2018) 7:766–73. doi: 10.21037/tau.2018.07.19
7. Kato S, Tanaka T, Kitamura H, Masumori N, Ito T, Kawaharada N, et al. Resection of the inferior vena cava for urological malignancies: single-center experience. *Int J Clin Oncol.* (2013) 18:905–9. doi: 10.1007/s10147-012-0473-x
8. Pecoraro A, Palumbo C, Knipper S, Rosiello G, Luzzago S, Tian Z, et al. Histologic subtype, tumor grade, tumor size, and race can accurately predict the probability of synchronous metastases in T2 renal cell carcinoma. *Clin Genitourin Cancer.* (2020) 18:e610–8. doi: 10.1016/j.clgc.2020.02.001
9. Zisman A, Pantuck AJ, Dorey F, et al. Improved prognostication of renal cell carcinoma using an integrated staging system. *J Clin Oncol.* (2001) 19:1649–57. doi: 10.1200/JCO.2001.19.6.1649
10. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: the SSIGN score. *J Urol.* (2002) 168:2395–400. doi: 10.1016/S0022-5347(05)64153-5
11. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, et al. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer.* (2003) 97:1663–71. doi: 10.1002/cncr.11234
12. Sheng X, Lu X, Wu J, Chen L, Cao H. A nomogram predicting the prognosis of renal cell carcinoma patients with lung metastases. *Biomed Res Int.* (2021) 2021:6627562. doi: 10.1155/2021/6627562
13. Wang K, Wu Z, Wang G, Shi H, Xie J, Yin L, et al. Survival nomogram for patients with bone metastatic renal cell carcinoma: a population-based study. *Int Braz J Urol.* (2021) 47:333–49. doi: 10.1590/s1677-5538.ibju.2020.0195
14. Sun M, De Velasco G, Brastianos PK, Aizer AA, Martin A, Moreira R, et al. The development of brain metastases in patients with renal cell carcinoma: epidemiologic trends, survival, and clinical risk factors using a population-based cohort. *Eur Urol Focus.* (2019) 5:474–81. doi: 10.1016/j.euf.2017.12.007
15. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Proc.* (2021). doi: 10.1049/ipr2.12222. [Epub ahead of print].
16. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics.* (2020) 9:274. doi: 10.3390/electronics9020274
17. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intel Hum Comput.* (2020). doi: 10.1007/s12652-020-01963-7. [Epub ahead of print].
18. Cronin KA, Ries LA, Edwards BK. The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. *Cancer.* (2014) 120(Suppl 23):3755–7. doi: 10.1002/cncr.29049
19. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* 2008;8:53. Published 2008 Nov 26. doi: 10.1186/1472-6947-8-53
20. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* (2015) 136:E359–86. doi: 10.1002/ijc.29210
21. Simard EP, Ward EM, Siegel R, Jemal A. Cancers with increasing incidence trends in the United States: 1999 through 2008 [published correction appears in CA Cancer J Clin. 2012 Jul-Aug;62(4):277]. *CA Cancer J Clin.* (2012) 62:118–28. doi: 10.3322/caac.20141
22. Tichet M, Prod'Homme V, Fenouille N, Ambrosetti D, A, Cerezo M, et al. Tumour-derived SPARC drives vascular permeability and extravasation through endothelial VCAM1 signalling to promote metastasis. *Nat Commun.* (2015) 6:6993. doi: 10.1038/ncomms7993
23. Cohen HT, McGovern FJ. Renal-cell carcinoma. *N Engl J Med.* (2005) 353:2477–90. doi: 10.1056/NEJMra043172
24. Franklin JR, Figlin R, Rauch J, Gitlitz B, Belldegrun A. Cytoreductive surgery in the management of metastatic renal cell carcinoma: the UCLA experience. *Semin Urol Oncol.* (1996) 14:230–6.
25. Choueiri TK, Escudier B, Powles T, Mainwaring PN, Rini BI, Donskov F, et al. Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomized, open-label, phase 3 trial. *Lancet Oncol.* (2016) 17:917–27. doi: 10.1016/S1470-2045(16)30107-3
26. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, et al. Epidemiology of renal cell carcinoma. *Eur Urol.* (2019) 75:74–84. doi: 10.1016/j.eururo.2018.08.036
27. Hutterer GC, Patard JJ, Perrotte P, Ionescu C, de La Taille A, Salomon L, et al. Patients with renal cell carcinoma nodal metastases can be accurately identified: external validation of a new nomogram. *Int J Cancer.* (2007) 121:2556–61. doi: 10.1002/ijc.23010
28. Capitanio U, Abdollah F, Matloob R, Suardi N, Castiglione F, Di Trapani E, et al. When to perform lymph node dissection in patients with renal cell carcinoma: a novel approach to the preoperative assessment of risk of lymph node invasion at surgery and of lymph node progression during follow-up. *BJU Int.* (2013) 112:E59–66. doi: 10.1111/bju.12125
29. Marconi L, de Bruijn R, van Werkhoven E, Beisland C, Fife K, Heidenreich A, et al. External validation of a predictive model of survival after cytoreductive nephrectomy for metastatic renal cell carcinoma. *World J Urol.* (2018) 36:1973–80. doi: 10.1007/s00345-018-2427-z
30. Bai X, Huang Q, Zuo P, Zhang X, Yuan J, Zhang X, et al. MRI radiomics-based nomogram for individualised prediction of synchronous distant metastasis in patients with clear cell renal cell carcinoma. *Eur Radiol.* (2021) 31:1029–42. doi: 10.1007/s00330-020-07184-y
31. Zhao Y, Liu G, Sun Q, Zhai G, Wu G, Li ZC. Validation of CT radiomics for prediction of distant metastasis after surgical resection in patients with clear cell renal cell carcinoma: exploring the underlying signaling pathways. *Eur Radiol.* (2021) 31:5032–40. doi: 10.1007/s00330-020-07590-2
32. Li Y, Chen P, Chen Z. A population-based study to predict distant metastasis in patients with renal cell carcinoma. *Ann Palliat Med.* (2021) 10:4273–88. doi: 10.21037/apm-20-2481
33. Hutterer GC, Patard JJ, Jeldres C, Perrotte P, de La Taille A, Salomon L, et al. Patients with distant metastases from renal cell carcinoma can be accurately identified: external validation of a new nomogram. *BJU Int.* (2008) 101:39–43. doi: 10.1111/j.1464-410X.2007.07170.x
34. Zastrow S, Phuong A, von Bar I, Novotny V, Hakenberg OW, Wirth MP. Primary tumor size in renal cell cancer in relation to the occurrence of synchronous metastatic disease. *Urol Int.* (2014) 92:462–7. doi: 10.1159/000356325
35. Pichler R, Compérat E, Klatte T, Pichler M, Loidl W, Lusuardi L, et al. Renal cell carcinoma with sarcomatoid features: finally new therapeutic hope? *Cancers.* (2019) 11:422. doi: 10.3390/cancers11030422

36. Taub DA, Miller DC, Cowan JA, Dimick JB, Montie JE, Wei JT. Impact of surgical volume on mortality and length of stay after nephrectomy. *Urology*. (2004) 63:862–7. doi: 10.1016/j.urology.2003.11.037

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Zhanghuang, Tan, Mi, Liu, Jin, Li, Zhang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 01 February 2022
doi: 10.3389/fpubh.2022.837506

A Nomogram for Predicting Cancer-Specific Survival of Osteosarcoma and Ewing's Sarcoma in Children: A SEER Database Analysis

*Jinkui Wang[1], Chenghao Zhanghuang[1,2], Xiaojun Tan[1,3], Tao Mi[1], Jiayan Liu[1], Liming Jin[1], Mujie Li[1], Zhaoxia Zhang[1] and Dawei He[1]**

[1] Department of Urology, Chongqing Key Laboratory of Children Urogenital Development and Tissue Engineering, Chongqing Key Laboratory of Pediatrics, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China, [2] Department of Urology, Kunming Children's Hospital (Children's Hospital Affiliated to Kunming Medical University), Yunnan Key Laboratory of Children's Major Disease Research, Kunming, China, [3] Department of Urology, Nanchong Central Hospital, the Second Clinical Medical College, North Sichuan Medical University, Nanchong, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Praveen Kumar,
VIT University, India
Rana Zeeshan Ahmad,
University of Sialkot, Pakistan

***Correspondence:**
Dawei He
hedawei@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 16 December 2021
Accepted: 03 January 2022
Published: 01 February 2022

Citation:
Wang J, Zhanghuang C, Tan X, Mi T, Liu J, Jin L, Li M, Zhang Z and He D (2022) A Nomogram for Predicting Cancer-Specific Survival of Osteosarcoma and Ewing's Sarcoma in Children: A SEER Database Analysis. Front. Public Health 10:837506. doi: 10.3389/fpubh.2022.837506

Background: Osteosarcoma (OSC) and Ewing's sarcoma (EWS) are children's most common primary bone tumors. The purpose of the study is to develop and validate a new nomogram to predict the cancer-specific survival (CSS) of childhood OSC and EWS.

Methods: The clinicopathological information of all children with OSC and EWS from 2004 to 2018 was downloaded from the Surveillance, Epidemiology, and End Results (SEER) database. Univariate and multivariate Cox regression analyses were used to screen children's independent risk factors for CSS. These risk factors were used to construct a nomogram to predict the CSS of children with OSC and EWS. A series of validation methods, including calibration plots, consistency index (C-index), and area under the receiver operating characteristic curve (AUC), were used to validate the accuracy and reliability of the prediction model. Decision curve analysis (DCA) was used to validate the clinical application efficacy of predictive models. All patients were divided into low- and high-risk groups based on the nomogram score. Kaplan-Meier curve and log-rank test were used to compare survival differences between the two groups.

Results: A total of 2059 children with OSC and EWS were included. All patients were randomly divided into training cohort 60% (N = 1215) and validation cohort 40% (N = 844). Univariate and multivariate analysis suggested that age, surgery, stage, primary site, tumor size, and histological type were independent risk factors. Nomograms were established based on these factors to predict 3-, 5-, and 8-years CSS of children with OSC and EWS. The calibration plots showed that the predicted value was highly consistent with the actual value. In the training cohort and validation cohort, the C-index was 0.729 (0.702–0.756) and 0.735 (0.702–0.768), respectively. The AUC of the training cohort and the validation cohort also showed similar results. The DCA showed that the nomogram had good clinical value.

Conclusion: We constructed a new nomogram to predict the CSS of OSC and EWS in children. This predictive model has good accuracy and reliability and can help doctors and patients develop clinical strategies.

Keywords: nomogram, cancer-specific survival, Osteosarcoma, Ewing's sarcoma, children

INTRODUCTION

Osteosarcoma (OSC) and Ewing sarcoma (EWS) are the most common primary malignant bone tumors in children and young people (1–3). OSC is the most common tumor in children and adolescence, and its peak incidence occurs in adolescence. The annual incidence rate for children under 10 years old is 1.7 per 100,000, while the yearly incidence rate for patients between 10 and 19 is 8.2 per 100,000 (4). Although various treatments are used to improve the prognosis, the survival rate of OSC is still shallow (5–7), the 5-year survival rate of patients with localized OSC can reach 65–70% (8). In comparison, the survival rate of patients with metastatic OSC is only 19–30% (9, 10). EWS is a malignant bone tumor commonly seen in children and adolescents. Its incidence is second only to OSC, accounting for 3% of solid malignant tumors in children (2, 3). In recent years, with the improvement of treatment technology, the 5-year survival rate of EWS has reached 60% to 70%, but the prognosis of patients with metastasis and recurrence is still poor (11).

Yang et al. (12) established a nomogram to predict the cancer-specific survival (CSS) and overall survival rate of OSC. It showed that age, stage, grade, surgery, primary site, and tumor size were independent risk factors for OSC. Lu et al. (13) established a nomogram to predict the risk of distant metastasis of OSC. We found that age, tumor location, tumor grade, T stage, and surgical method were risk factors for metastasis and suggested a worse prognosis. Zhou et al. (14) established a nomogram to predict the overall survival of EWS, revealing that age, N stage and bone metastasis are independent risk factors for prognosis. However, as far as we know, no nomogram has been established for children with OSC and EWS to predict tumor-specific survival.

At present, artificial intelligence has been widely used in human health. Dhanamjayulu et al. (15) used machine learning to identify malnourished people. Gadekallu et al. (16) used a neural network to predict retinopathy in diabetic patients. Reddy et al. (17) used an artificial intelligence algorithm to diagnose heart disease. Abbas et al. (18) used a new algorithm to monitor breast cancer. Mubashar et al. (19) use artificial intelligence to manage personal health records.

The nomogram is a way to estimate the occurrence of a given event by generating corresponding values from the clinical-pathological information of the patient. At present, nomograms have been widely used in the clinical prediction of liver cancer, lung cancer, and breast cancer (20–22). As the two most common bone malignancies in children, OSC and EWS seriously endanger children's health. Revealing the prognostic factors of childhood OSC and EWS can help doctors and patients choose appropriate treatment measures, which is conducive to prolonging the survival time of patients and improving the quality of life. In other words, accurately predicting the prognosis of malignant tumors can help patients and doctors formulate treatment plans and follow-up strategies.

However, the prediction models for OSC and EWS in children are still not well established. No studies have precisely predicted CSS in pediatric OSC and EWS patients. Bone tumor in children is an essential factor endangering children's health. Accurate prediction of the survival of bone tumors in children can enhance the population's health and improve people's overall quality of life and life span. Therefore, it is necessary to construct a nomogram to predict the survival prognosis of childhood OSC and EWS.

Abbreviations: OSC, Osteosarcoma; EWS, Ewing's sarcoma; SEER, the Public Surveillance, Epidemiology, and End Results; CSS, cancer-specific survival; C-index, concordance index; AUC, area under the receiver operating characteristic curve; DCA, decision curve analysis.

PATIENTS AND METHODS

Data Source and Data Extraction

Patient data are downloaded from the Surveillance, Epidemiology, and End Results (SEER) project of the National Cancer Institute. From 2004 to 2018, all children under 18 diagnosed with OSC and EWS were collected. The SEER database is a cancer database in the United States, covering about 28% of Americans, and contains 18 cancer registries (23). The SEER database discloses patient demographic information, clinicopathological characteristics, and survival status. The clinicopathological information we used is public and anonymous, so our study does not require ethical approval and patient consent. Our study method complies with the regulations of the SEER database.

We collected patient demographic information (age, sex, race), tumor characteristics (tumor grade, size, primary location, pathological type), treatment (surgery, radiotherapy, chemotherapy), follow-up information (survival status, survival time). The selection criteria are: (1) age less than 18 years old; (2) diagnosed as OSC and EWS. The exclusion criteria are: (1) the surgical method is unknown; (2) non-primary tumor; (3) the tumor size is unknown; (4) The location of the tumor is not precise; (5) Survival time less than 1 month. The flowchart of patient screening is shown in **Figure 1**.

The race of all children was divided into white, black, and other (American Indian/AK Indian, Asian/Pacific Islander). The tumor grade was split into I (well-differentiated), II (medium differentiated), III (poorly differentiated), IV (undifferentiated). According to the SEER surgery code, the surgery methods were divided into four groups: no surgery, partial resection, radical excision, and amputation. The primary site of the tumor was divided into the limb, cranial, spine, thoracic, and pelvic.

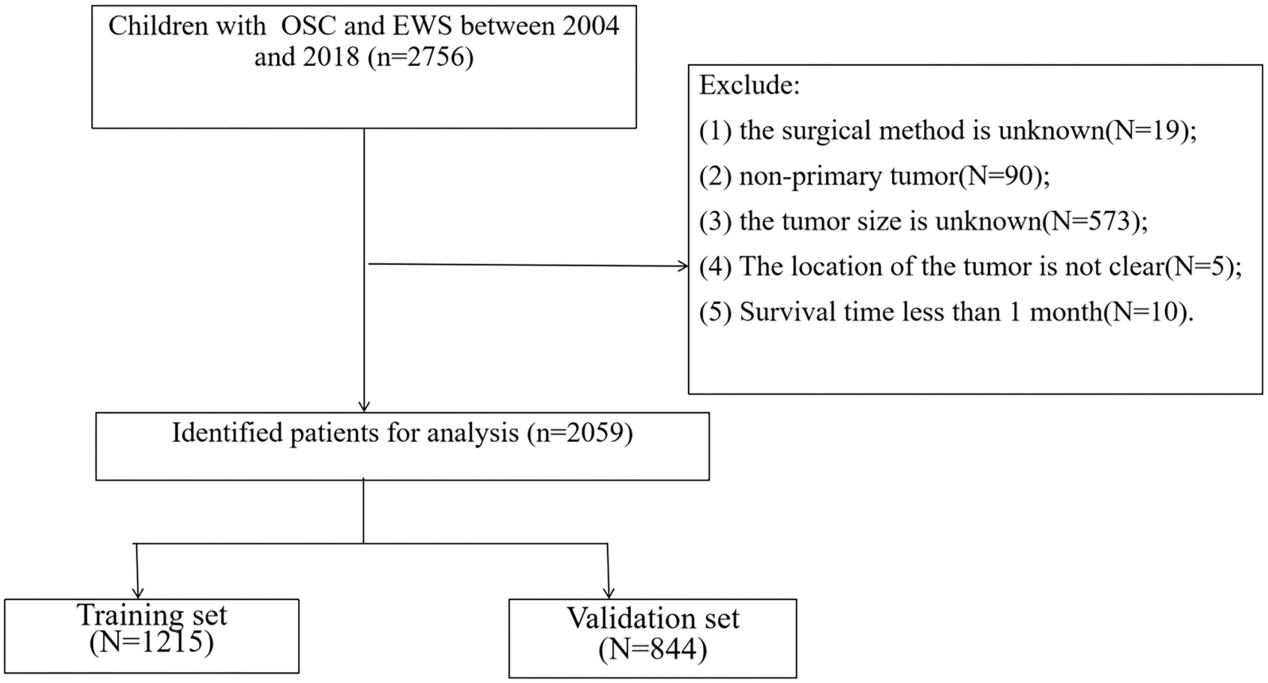


FIGURE 1 | The flowchart of including and dividing patients.

Univariate and Multivariate Cox Regression Analysis

All patients were randomly divided into a training set (60%) and a validation set (40%). Univariate Cox regression screened out related prognostic factors. Then multivariate Cox regression analysis was used to analyse the independent risk factors of the training set. And we recorded the hazard ratio (HR) and 95% confidence interval (CI) simultaneously.

Nomogram Construction

Based on the independent risk factors obtained by univariate and multivariate Cox regression, a nomogram was constructed to predict the 3-, 5-, 8-year CSS of children with OSC and EWS. Each variable is distributed on the nomogram according to its weight to get different lines, and the points of each variable correspond to a point. The sum of the points of all the variables in the nomogram can equal an overall points, thus obtaining survival rates at different points in time.

Nomogram Validation

Then a series of validation methods were used, including the consistency index (C-index), the calibration curve, and the area under the receiver operating characteristic curve (AUC). The accuracy of the nomogram was tested mainly by the calibration curve, which was used to compare the relationship between the observed value and the actual value through 1,000 bootstrap sampling. The results of the calibration curve are dotted in the **Figure 3**. If they are close to the diagonal, the model has good accuracy. Both the C-index and the AUC are used to assess the discrimination of the model.

Clinical Utility

Decision curve analysis (DCA) is a new algorithm to evaluate the clinical value of the model based on the net benefit under each risk threshold (24). DCA was used to assess the clinical potential application value of the new nomogram. Subsequently, we divided the patients into a low-risk group and a high-risk group through the nomogram score of each patient. The Kaplan-Meier curve and log-rank test were used to compare the survival differences of different risk groups. At the same time, we analyzed the surgical methods of patients in other risk groups.

Statistical Analysis

All statistics were performed using SPSS23.0 (IBM, Chicago, IL, USA) and R software (version 3.4.1; http://www.Rproject.org). Age and tumor size were continuous variables and did not follow a normal distribution; median and the inter-quartile range was used for description. All possible prognostic factors were confirmed using univariate Cox regression. Then, the relevant variables were included in the multivariate analysis and the stepwise regression method was used to select the best Cox regression model. A $p < 0.05$ was considered statistically significant.

RESULTS

Clinical Features

A total of 2,059 children with OSC and EWS were included in our study. They were divided into a training set (N = 1215) and a validation set (N = 844). The clinicopathological characteristics are shown in **Table 1**. There was no significant difference between the training set and the validation set. The median age of all children was 13 years old (interquartile range, 10–15 years old), including 1,179 males (57.3%) and 1,624 whites (78.9%). One thousand three hundred eighty-five cases (67.3%) were OSC, and 674 cases (32.7%) were EWS. Children whose primary tumors were located in limb, cranial, spine, thoracic and pelvic were 1,576 cases (76.5%), 93 cases (4.52%), 56 cases (2.72%), 118 cases (5.73%), and 216 cases (10.5%), respectively. Children with tumor grades I, II, III, and IV was 23 (1.12%), 48 (2.33%), 419 (20.3%), and 747 (36.3%), respectively. There were 673 cases (32.7%), 842 cases (40.9%), and 531 cases (25.8%) of children whose tumor stages were localized, regional and distant, respectively. The median tumor size was 90 mm (interquartile range, 62–125 mm). Most children have performed chemotherapy (1,978, 96.1%), and most have not completed radiotherapy (1,674, 81.3%). Partial resection, radical excision and amputation were performed in 230 cases (11.2%), 1,166 cases (56.6%), and 342 cases (16.6%). A total of 474 patients (23.0%) had distant metastases.

Univariate and Multivariate Cox Regression Analysis

We first established a univariate Cox regression model to screen possible prognostic factors. Then we established a Cox multivariate model to identify independent risk factors affecting OS in children. Univariate and multivariate Cox regression results are shown in **Table 2**. Finally, we found age (HR 1.030; 95% CI 1.006–1.055), tumor size (HR 1.002; 95% CI 1.001–1.003), histological type (HR 0.563; 95% CI 0.434–0.73), surgery, stage, and primary site were independent risk factors. In other words, these six factors can be used to establish a nomogram to predict the CSS of OSC and EWS in children.

Nomogram Construction for 3-Year, 5-Year, and 8-Year CCS

The six independent risk factors above were used to construct a nomogram to predict CSS in children with OSC and EWS (**Figure 2**). The nomogram accurately listed the impact of each factor on CSS. We found that tumor size was the most significant influencing factor, and a large tumor meant a higher risk of death. The second was the staging of tumors. There was no doubt that distantly metastatic tumors had a higher risk of death than in localized or regional tumors. The location of the primary tumor was also a significant risk factor. We found that tumors in the spine, pelvis, and thorax have a significantly higher risk of death than tumors in the limbs and skull. In addition, the children with radical resection of the tumor, the younger children and the children with EWS had the lowest mortality.

Validation of the Nomogram

The C-index of the training set and the validation set were 0.729 (95% CI 0.702–0.756) and 0.735 (95% CI 0.702–0.768), respectively. This showed that the nomogram has good discrimination. The calibration curves of the training set and the validation set showed that the predicted value of the nomogram is highly consistent with the observed value (**Figures 3A–F**), which proved that the nomogram has good accuracy. In the training set, the 3-, 5-, 8-year AUC of the nomogram were 74.7 (95% CI 71–78.4), 72.5 (95% CI 68.9–76.1), 68.7 (95% CI 64.5–73.0), respectively. In the validation set, the 3-, 5-, 8-year AUC of the nomogram were 73.6 (95% CI 69.0–78.2), 71.2 (95% CI 66.5–75.9), and 72.3 (95% CI 67.3–77.3), respectively (**Figures 4A,B**). The AUC results once again proved the accuracy and discrimination of the nomogram. These validations showed that the nomogram was at least 70% accurate, especially for predicting medium-term survival.

Clinical Application of the Nomogram

The DCA of the training set and the validation set showed that the clinical value of the nomogram is higher than that of the tumor stage (**Figures 5A,B**). According to the score of the nomogram, the patients were divided into two groups: low-risk group (total score $<$75.46) and high-risk group (total score $\geq$75.46). The Kaplan-Meier curve showed that high-risk patients have lower survival rates than low-risk patients (**Figures 6A,B**). In the total set, the low-risk group's 3-year, 5-year, and 8-year CSS rates were 92.2, 89.7, and 81.5%, the high-risk group were 75.7, 67.5, and 63.3%, respectively. In addition, we analyzed the impact of surgical methods in different risk groups on the survival probability of patients. In the low-risk group, radical resection was the primary surgical method, and there was no significant difference in the survival rate of various surgical methods (**Figure 7A**). In the high-risk group, patients with radical resection had the highest survival rate, followed by partial resection, amputation and no surgery (**Figure 7B**).

DISCUSSION

OSC is the most common malignant bone tumor in children. The application of neoadjuvant chemotherapy has increased the survival rate of patients with OSC from 15 to 17% of surgical treatment alone to 70% (25). However, the survival rate of patients with OSC has not improved significantly in the past 30 years. As the second most common bone tumor in children, EWS has a lower mortality rate than OSC, but metastatic EWS is still a fundamental cause of harm to children's health (26). Because it is a highly aggressive cancer, it can metastasise to bone marrow, lung and other tissues in the early stage of the disease (27–29): children's malignant bone tumors, especially OSC, hurt patients. Accurately assessing prognostic factors can improve the prognosis and help clinicians effectively evaluate survival and make treatment decisions.

We collected 2,059 children with OSC and EWS from the SEER database. The results showed that the survival rate of children with EWS was significantly higher than that of OSC. More whites than blacks and other races and more boys than girls

TABLE 1 | Clinicopathological characteristics of children with bone tumor.

	Total (*N* = 2,059)	Training cohort (*N* = 1,215)	Validation cohort (N = 844)	*p*
Age (median [IQR])	13 (10, 15)	13 (10, 15)	13 (10, 15)	0.6892
Race				0.717
White	1,624 (78.9%)	955 (78.6%)	669 (79.3%)	
Black	233 (11.3%)	143 (11.8%)	90 (10.7%)	
Other	202 (9.81%)	117 (9.63%)	85 (10.1%)	
Sex				0.020
Male	1,179 (57.3%)	722 (59.4%)	457 (54.1%)	
Female	880 (42.7%)	493 (40.6%)	387 (45.9%)	
Grade				0.933
I	23 (1.12%)	14 (1.15%)	9 (1.07%)	
II	48 (2.33%)	27 (2.22%)	21 (2.49%)	
III	419 (20.3%)	248 (20.4%)	171 (20.3%)	
IV	747 (36.3%)	433 (35.6%)	314 (37.2%)	
Unknown	822 (39.9%)	493 (40.6%)	329 (39.0%)	
Stage				0.591
Localized	673 (32.7%)	398 (32.8%)	275 (32.6%)	
Regional	842 (40.9%)	484 (39.8%)	358 (42.4%)	
Distant	531 (25.8%)	325 (26.7%)	206 (24.4%)	
Unstaged	13 (0.63%)	8 (0.66%)	5 (0.59%)	
Year of diagnosis				0.651
2004–2010	860 (41.8%)	502 (41.3%)	358 (42.4%)	
2010–2018	1,199 (58.2%)	713 (58.7%)	486 (57.6%)	
Primary site				0.713
Limb	1,576 (76.5%)	926 (76.2%)	650 (77.0%)	
Cranial	93 (4.52%)	55 (4.53%)	38 (4.50%)	
Spine	56 (2.72%)	35 (2.88%)	21 (2.49%)	
Thoracic	118 (5.73%)	76 (6.26%)	42 (4.98%)	
Pelvic	216 (10.5%)	123 (10.1%)	93 (11.0%)	
Histologic type				0.832
Osteogenic sarcoma	1,385 (67.3%)	820 (67.5%)	565 (66.9%)	
Ewings sarcoma	674 (32.7%)	395 (32.5%)	279 (33.1%)	
Laterality				0.789
Left	952 (46.2%)	563 (46.3%)	389 (46.1%)	
Right	901 (43.8%)	535 (44.0%)	366 (43.4%)	
Not a paired site	206 (10.0%)	117 (9.63%)	89 (10.5%)	
Chemotherapy				0.189
No/Unknown	81 (3.93%)	54 (4.44%)	27 (3.20%)	
Yes	1,978 (96.1%)	1,161 (95.6%)	817 (96.8%)	
Radiation				0.937
No/Unknown	1,674 (81.3%)	989 (81.4%)	685 (81.2%)	
Yes	385 (18.7%)	226 (18.6%)	159 (18.8%)	
Surgery				0.926
No	321 (15.6%)	193 (15.9%)	128 (15.2%)	
Partial resection	230 (11.2%)	136 (11.2%)	94 (11.1%)	
Radical excision	1,166 (56.6%)	681 (56.0%)	485 (57.5%)	
Amputation	342 (16.6%)	205 (16.9%)	137 (16.2%)	
Tumor size (median [IQR])	90 (62, 125)	90 (63, 125)	88.5 (61.75, 125.25)	0.7592
Metastasis				0.469
No/Unknown	1,585 (77.0%)	928 (76.4%)	657 (77.8%)	
Yes	474 (23.0%)	287 (23.6%)	187 (22.2%)	

TABLE 2 | Univariate and multivariate analyses of CSS in bone tumor.

	Univariate			Multivariate		
	HR	**95% CI**	***p***	**HR**	**95% CI**	***p***
Age	1.06	1.03–1.09	<0.0001	1.030	1.006–1.055	0.015
Sex						
Male	Reference					
Female	0.87	0.72–1.07	0.19			
Year of diagnosis						
2004–2010	Reference					
2011–2018	0.53	0.43–0.64	<0.0001			
Race						
White	Reference					
Black	1.13	0.83–1.53	0.45			
Other[a]	0.94	0.67–1.31	0.71			
Primary site						
Limb	Reference					
Cranial	0.45	0.24–0.83	0.01	0.772	0.425-1.401	0.394
Spine	1.21	0.67–2.19	0.52	1.760	1.021–3.033	0.042
Thoracic	1.03	0.67–1.59	0.88	1.232	0.814–1.864	0.324
Pelvic	2.12	1.58–2.85	<0.0001	1.396	1.045–1.863	0.024
Grade						
I	Reference					
II	0.48	0.14–1.65	0.25			
III	0.85	0.33–2.22	0.75			
IV	1.17	0.45–3	0.75			
Unknown	0.97	0.38–2.5	0.95			
Histologic type						
Osteogenic sarcoma	Reference			Reference		
Ewings sarcoma	0.89	0.72–1.1	0.28	0.563	0.434–0.73	<0.0001
Laterality						
Left	Reference					
Right	0.91	0.74–1.11	0.35			
Not a paired site	0.99	0.7–1.38	0.93			
Stage						
Localized	Reference			Reference		
Regional	2.3	1.74–3.03	<0.0001	1.743	1.348–2.255	<0.0001
Distant	5.96	4.48–7.93	<0.0001	4.400	3.396–5.701	<0.0001
Unstaged	1.27	0.28–5.85	0.75	1.309	0.319–5.377	0.709
Surgery						
No	Reference			Reference		
Partial resection	0.55	0.37–0.8	<0.0001	0.707	0.499–1.002	0.051
Radical excision	0.46	0.35–0.6	<0.0001	0.503	0.382–0.663	<0.0001
Amputation	0.87	0.63–1.2	0.40	0.742	0.545–1.01	0.058
Radiation						
No/Unknown	Reference					
Yes	1.63	1.28–2.07	<0.0001			
Chemotherapy						
No/Unknown	Reference					
Yes	2.31	1.21–4.4	0.01			
Tumor size	1	1–1.01	<0.0001	1.002	1.001–1.003	0.006
Metastasis						
No	Reference					
Yes	3.58	2.88–4.46	3.58			

[a]*Other: American Indian/AK Native, Asian/Pacific Islander.*

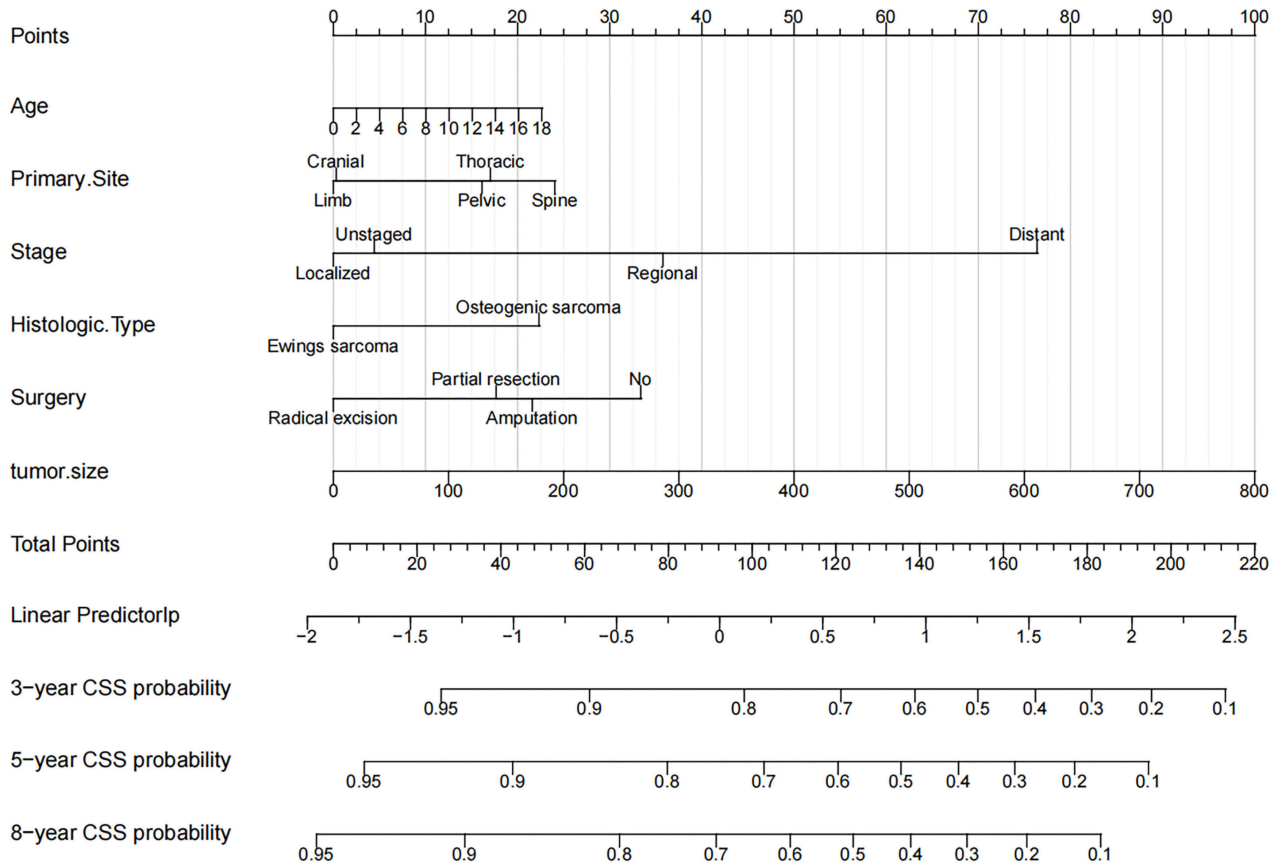


FIGURE 2 | Nomogram for 3-, 5-, and 8-year CSS of children with OSC and EWS.

in all children. The median age of all children was 13 years old, and the peak incidence was 12–15 years old. This study found six independent risk factors through univariate and multivariate Cox regression models, including age, surgery, stage, primary site, tumor size, and histological type. The nomogram includes these factors to establish a predictive model to predict the 3-, 5-, 8-year CSS of childhood OSC and EWS. We first use a nomogram to predict and validate CSS in childhood OSC and EWS. This nomogram has good accuracy and reliability and is of great significance to clinical patients. The nomogram is user-friendly and can be applied to everyone with simple learning.

Previous studies have found prognostic factors for OSC and EWS. However, no study has used these influencing factors to construct a prediction model for survival prediction of children with a bone tumor. Although the influencing factors we identified have previously been shown to be independent risk factors, we integrated these risk factors to construct a nomogram predicting long-term survival in children with OSC and EWS.

Previous studies have shown that various prognostic factors (age, tumor location, type of surgery, local recurrence) will affect the survival of patients (30–32). Our study found that age is an independent risk factor for malignant bone tumors in children, and the survival rate of older children is lower, the same as previous research results (33, 34). Lee et al. (35) found that fewer adult patients receive chemotherapy, which may also be a reason. In addition, we found that the tumor's location near the central axis (spine, pelvis, thorax) has a significantly higher risk of death than the limbs. Like previous studies, axial tumor location predicts the worst prognosis (36, 37). Wan et al. (38) also found that the survival rate of EWS of the spine and pelvis was lower. Because the tumors in the axial position are easier to infiltrate vital organs and metastases far away than the tumors of the limbs. Oberlin et al. (39) believe that bone tumors of the limbs can be removed by surgery, but it is more difficult to remove the axial bone's tumors. Similarly, as in a previous report (40), our study found that tumor size is also a significant risk factor. Large tumors are prone to metastasis, indicating a worse prognosis.

In addition, surgical treatment of bone tumors in children, especially in extremity tumors, is still an important issue. With the widespread development of neoadjuvant chemotherapy and standardized surgery, the survival rate of patients with malignant bone tumors has improved significantly, and limb salvage has become the primary surgical method in treating malignant bone tumors. This treatment model can preserve the function of the limbs and joints and effectively reduce the metastasis and recurrence of OSC (41). Whether it is the low-risk or high-risk group, radical salvage resection has the highest survival rate in our study. However, overall,

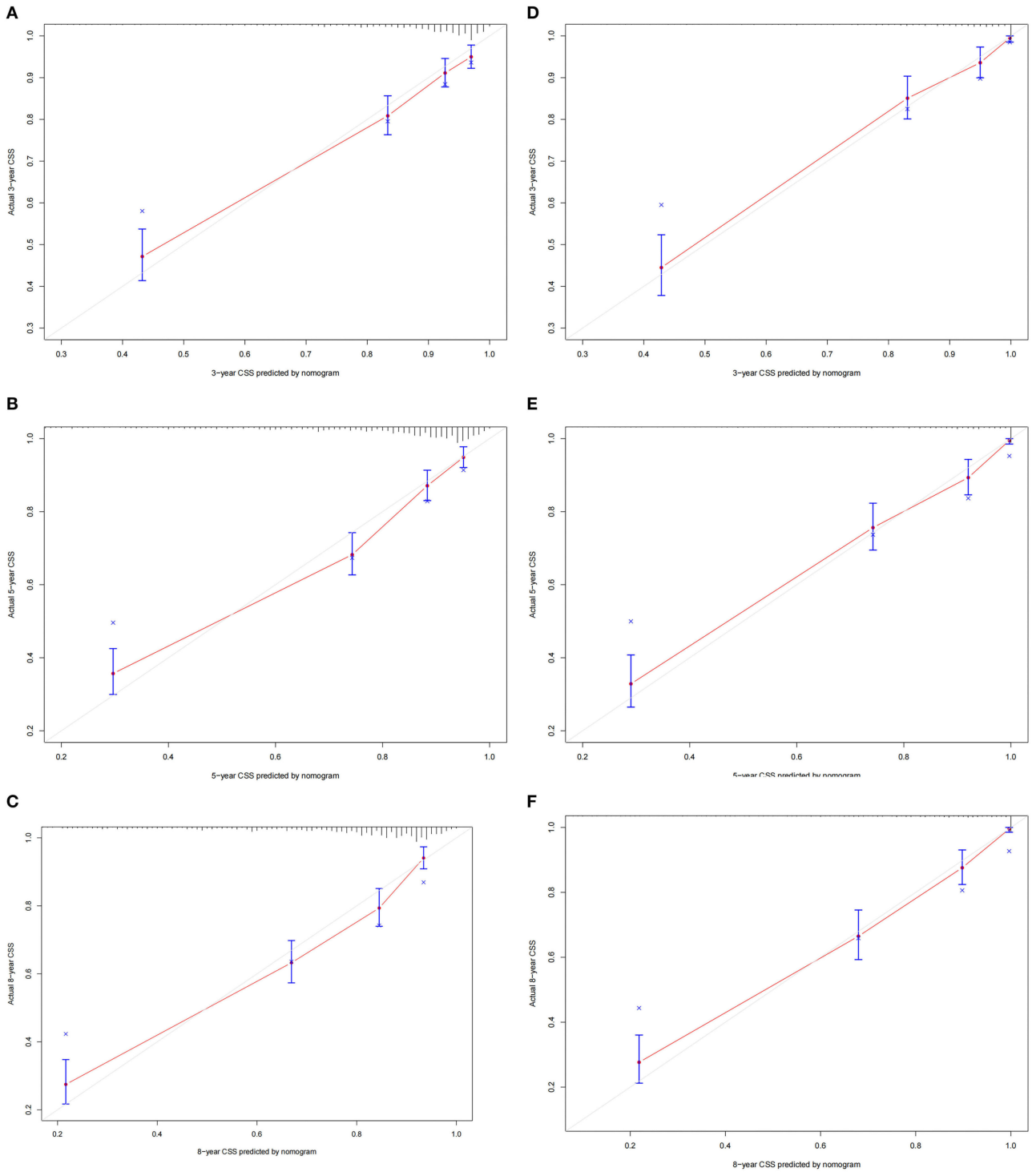


FIGURE 3 | Calibration curve of the nomogram. **(A–C)** For 3-, 5-, and 8-year CSS in the training set; **(D–F)** for 3-, 5-, and 8-year in the validation set.

patients with surgery have a higher survival rate than those without surgery.

Chemotherapy and surgery are still the main treatments for OSC (1). In our study, almost all patients received chemotherapy. In addition, radiotherapy seems to be controversial for treating bone tumors in children. Because surgery is still a radical cure for malignant bone tumors, it will only be considered if cancer cannot be removed entirely (42). For children, radiotherapy

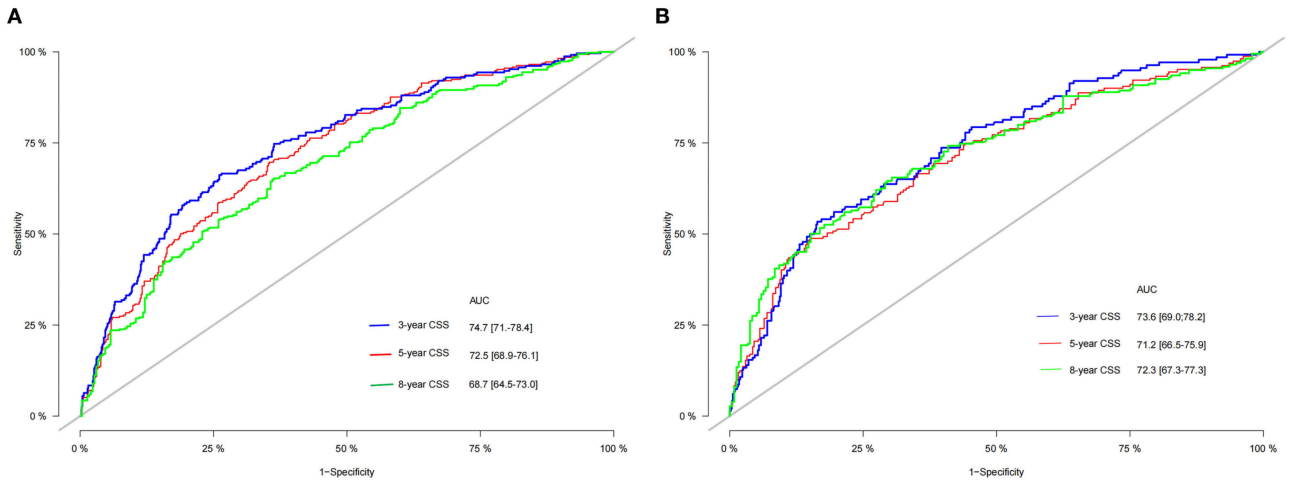


FIGURE 4 | The ROC of 3-, 5-, and 8-year of the training **(A)** and validation **(B)** sets.

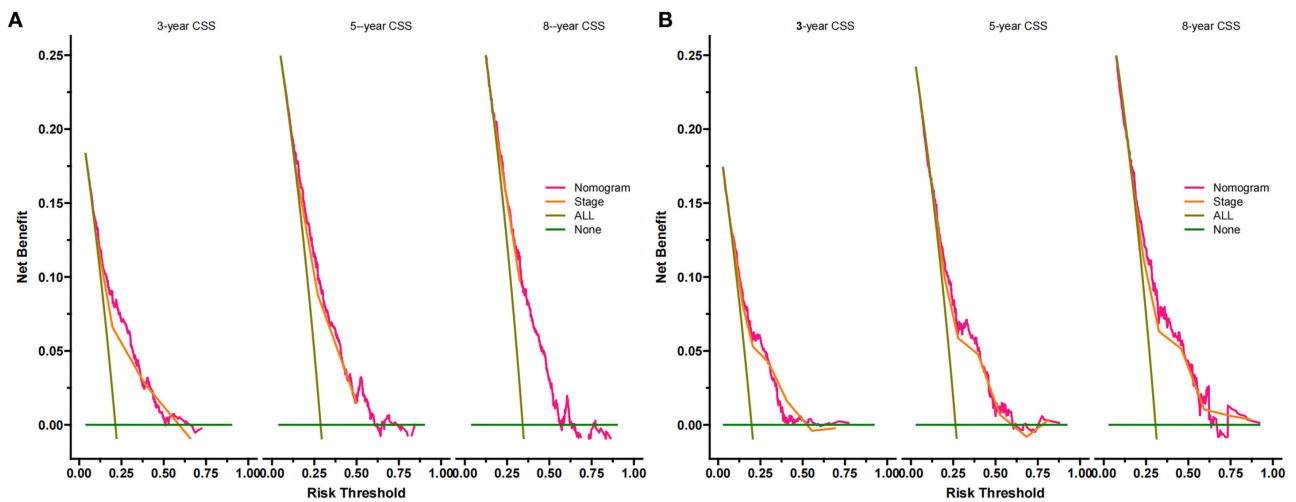


FIGURE 5 | Decision curves of the nomogram predicting CSS in the training set **(A)** and the validation set **(B)**. The x-axis is the threshold probability, and the y-axis is the net benefit. The green line indicates that no patients have died, and the dark green line indicates that all patients have died. When the threshold probability is between 10 and 50%, the net benefit of the model exceeds all deaths or no deaths.

may cause developmental delay of bones or organs. Even if it has a therapeutic effect on tumors, it will also seriously affect the quality of life in the future. Our research also found that only a few patients received radiotherapy, similar to previous studies (43).

It is worth mentioning that our analysis found that sex, race, and tumor grade are not independent risk factors. Previous studies have found that they are indeed related to the patient's survival prognosis (44–46). Regarding the grade of the tumor, it may be related to too many cases of unclear grades. Whether race and sex factors affect, the prognosis needs further research.

However, our research still has certain limitations. First, because our study is a retrospective case study based on the SEER database, some possible variables such as surgical margins, tumor recurrence, genetic factors, etc., are not available. However, we have included essential variables such as tumor size, tumor location, pathological type and other vital elements that determine the prognosis and will not cause devastating deviations. Second, we included patients from 2004 to 2018. With the improvement of treatment methods, the survival rate of patients in different years should be different. However, we stratified by the year of diagnosis and found no significant difference, indicating that it did not affect the results. Finally, all the data used in our study were downloaded from the SEER database, and the constructed nomogram lacks external data to validate. Therefore, it is necessary to use external validation further to test the accuracy and reliability of the prediction model. Next, we will conduct further prospective studies in our hospital to externally validate the accuracy of this prediction model further to promote the clinical application.

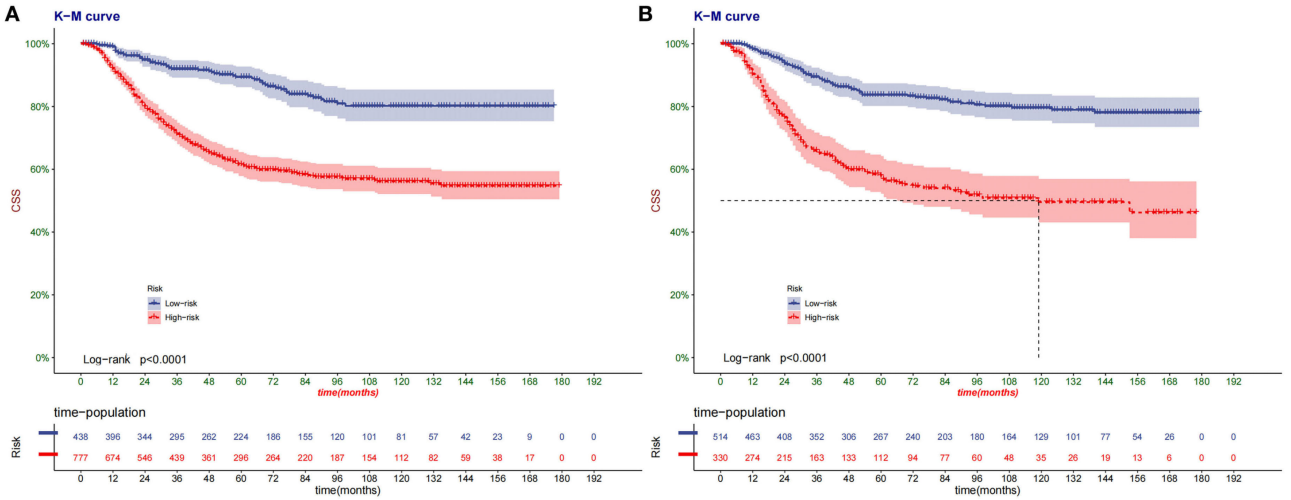


FIGURE 6 | Kaplan–Meier curves of CSS for children in the low- and high-risk groups in the training set **(A)** and validation set **(B)**.

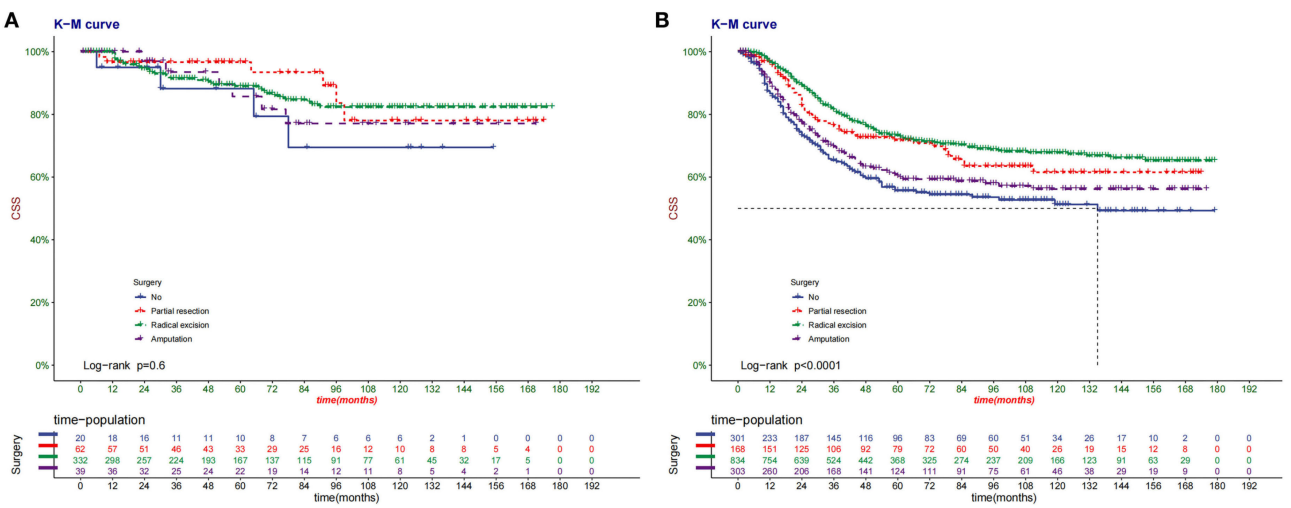


FIGURE 7 | Kaplan–Meier curves of CSS for children with different surgery in the low- **(A)** and high-risk **(B)** groups in total set.

CONCLUSION

We constructed a new nomogram to predict the CSS of children with OSC and EWS. In addition, we found that age, surgery, stage, primary site, tumor size, and histological type are important risk factors affecting children with OSC and EWS. The discovery of risk factors and the Construction of nomograms can help doctors accurately grasp the prognosis of patients, answer patient consultations and help patients make clinical decisions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov/.

AUTHOR CONTRIBUTIONS

JW and CZ designed the study. CZ, JW, JL, ML, LJ, and XT collected and analyzed the data. JW drafted the initial manuscript. CZ, TM, JL, ZZ, and DH revised the article critically. CZ, JL, DH, ML, and XT reviewed and edited the article. All authors approved the final manuscript.

FUNDING

This work was funded by Special Key Project of Chongqing Technology Innovation and Application Development (No. Cstc2019jscx-tjsbX0003).

REFERENCES

1. Isakoff MS, Bielack SS, Meltzer P, Gorlick R. Osteosarcoma: current treatment and a collaborative pathway to success. *J Clin Oncol.* (2015) 33:3029–35. doi: 10.1200/JCO.2014.59.4895
2. Orr WS, Denbo JW, Billups CA, Wu J, Navid F, Rao BN, et al. Analysis of prognostic factors in extraosseous Ewing sarcoma family of tumors: review of St. Jude Children's Research Hospital experience. *Ann Surg Oncol.* (2012) 19:3816–22. doi: 10.1245/s10434-012-2458-4
3. Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin.* (2014) 64:83–103. doi: 10.3322/caac.21219
4. Nagarajan R, Weigel BJ, Thompson RC, Perentesis JP. Osteosarcoma in the first decade of life. *Med Pediatr Oncol.* (2003) 41:480–3. doi: 10.1002/mpo.10403
5. Marina NM, Smeland S, Bielack SS, Bernstein M, Jovic G, Krailo MD, et al. Comparison of MAPIE versus MAP in patients with a poor response to preoperative chemotherapy for newly diagnosed high-grade osteosarcoma (EURAMOS-1): an open-label, international, randomised controlled trial. *Lancet Oncol.* (2016) 17:1396–408. doi: 10.1016/S1470-2045(16)30214-5
6. Martin-Broto J, Redondo A, Valverde C, Vaz MA, Mora J, Garcia Del Muro X, et al. Gemcitabine plus sirolimus for relapsed and progressing osteosarcoma patients after standard chemotherapy: a multicenter, single-arm phase II trial of Spanish Group for Research on Sarcoma (GEIS). *Ann Oncol.* (2017) 28:2994–9. doi: 10.1093/annonc/mdx536
7. Smrke A, Anderson PM, Gulia A, Gennatas S, Huang PH, Jones RL. Future directions in the treatment of osteosarcoma. *Cells.* (2021) 10:172. doi: 10.3390/cells10010172
8. Meyers PA, Schwartz CL, Krailo M, Kleinerman ES, Betcher D, Bernstein ML, et al. Osteosarcoma: a randomized, prospective trial of the addition of ifosfamide and/or muramyl tripeptide to cisplatin, doxorubicin, and high-dose methotrexate. *J Clin Oncol.* (2005) 23:2004–11. doi: 10.1200/JCO.2005.06.031
9. Meyers PA, Heller G, Healey JH, Huvos A, Applewhite A, Sun M, et al. Osteogenic sarcoma with clinically detectable metastasis at initial presentation. *J Clin Oncol.* (1993) 11:449–53. doi: 10.1200/JCO.1993.11.3.449
10. Kager L, Zoubek A, Pötschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols. *J Clin Oncol.* (2003) 21:2011–8. doi: 10.1200/JCO.2003.08.132
11. Leavey PJ, Collier AB. Ewing sarcoma: prognostic criteria, outcomes and future treatment. *Expert Rev Anticancer Ther.* (2008) 8:617–24. doi: 10.1586/14737140.8.4.617
12. Yang QK, Lai QY, Wang Y, Wang Y, Yao ZX, Zhang XJ. Establishment and validation of prognostic nomograms to predict overall survival and cancer-specific survival for patients with Osteosarcoma. *Neoplasma.* (2021) 68:434–46. doi: 10.4149/neo_2020_200617N639
13. Lu S, Wang Y, Liu G, Wang L, Wu P, Li Y, et al. Construction and validation of nomogram to predict distant metastasis in osteosarcoma: a retrospective study. *J Orthop Surg Res.* (2021) 16:231. doi: 10.1186/s13018-021-02376-8
14. Zhou Z, Wang J, Fang L, Ma J, Guo M. A nomogram for predicting overall survival in patients with Ewing sarcoma: a SEER-based study. *BMC Musculoskelet Disord.* (2020) 21:737. doi: 10.1186/s12891-020-03706-3
15. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021). doi: 10.1049/ipr2.12222
16. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intell Human Comput.* (2020) 2020:1–14. doi: 10.1007/s12652-020-01963-7
17. Reddy GT, Reddy MPK, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell.* (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
18. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci.* (2021) 7:e390. doi: 10.7717/peerj-cs.390
19. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm. *J Circuits Syst Comput.* (2021) 2021:2250010. doi: 10.1142/S0218126622500104
20. van den Hoven I, van Klaveren D, Verheuvel NC, van la Parra RFD, Voogd AC, de Roos WK, et al. Predicting the extent of nodal involvement for node positive breast cancer patients: development and validation of a novel tool. *J Surg Oncol.* (2019) 120:578–86. doi: 10.1002/jso.25644
21. Zhang Y, Zheng D, Xie J, Li Y, Wang Y, Li C, et al. Development and validation of web-based nomograms to precisely predict conditional risk of site-specific recurrence for patients with completely resected non-small cell lung cancer: a multiinstitutional study. *Chest.* (2018) 154:501–11. doi: 10.1016/j.chest.2018.04.040
22. Hu T, Wang S, Huang L, Wang J, Shi D, Li Y, et al. A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. *Eur Radiol.* (2019) 29:439–49. doi: 10.1007/s00330-018-5539-3
23. Cronin KA, Ries LA. Edwards BK. The Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. *Cancer.* (2014) 120 Suppl 23:3755–7. doi: 10.1002/cncr.29049
24. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak.* (2008) 8:53. doi: 10.1186/1472-6947-8-53
25. Luetke A, Meyers PA, Lewis I, Juergens H. Osteosarcoma treatment - where do we stand? A state of the art review. *Cancer Treat Rev.* (2014) 40:523–32. doi: 10.1016/j.ctrv.2013.11.006
26. Grünewald TGP, Cidre-Aranaz F, Surdez D, Tomazou EM, de Álava E, Kovar H, et al. Ewing sarcoma. *Nat Rev Dis Primers.* (2018) 4:5. doi: 10.1038/s41572-018-0003-x
27. Uyeturk U, Helvaci K, Demirci A, Sonmez OU, Turker I, Afsar CU, et al. Clinical outcomes and prognostic factors of adult's Ewing sarcoma family of tumors: single center experience. *Contemp Oncol.* (2016) 20:141–6. doi: 10.5114/wo.2016.58487
28. Delattre O, Zucman J, Melot T, Garau XS, Zucker JM, Lenoir GM, et al. The Ewing family of tumors—a subgroup of small-round-cell tumors defined by specific chimeric transcripts. *N Engl J Med.* (1994) 331:294–9. doi: 10.1056/NEJM199408043310503
29. Teicher BA, Bagley RG, Rouleau C, Kruger A, Ren Y, Kurtzberg L. Characteristics of human Ewing/PNET sarcoma models. *Ann Saudi Med.* (2011) 31:174–82. doi: 10.4103/0256-4947.78206
30. Bacci G, Longhi A, Versari M, Mercuri M, Briccoli A, Picci P. Prognostic factors for Osteosarcoma of the extremity treated with neoadjuvant chemotherapy: 15-year experience in 789 patients treated at a single institution. *Cancer.* (2006) 106:1154–61. doi: 10.1002/cncr.21724
31. Kim MS, Lee SY, Lee TR, Cho WH, Song WS, Koh JS, et al. Prognostic nomogram for predicting the 5-year probability of developing metastasis after neo-adjuvant chemotherapy and definitive surgery for AJCC stage II extremity osteosarcoma. *Ann Oncol.* (2009) 20:955–60. doi: 10.1093/annonc/mdn723
32. Grimer RJ, Taminiau AM, Cannon SR, Surgical Subcommitte of the European Osteosarcoma Intergroup. Surgical outcomes in Osteosarcoma. *J Bone Joint Surg Br.* (2002) 84:395–400. doi: 10.1302/0301-620x.84b3.12019
33. Bacci G, Ferrari S, Mercuri M, Bertoni F, Picci P, Manfrini M, et al. Predictive factors for local recurrence in osteosarcoma: 540 patients with extremity tumors followed for minimum 25 years after neoadjuvant chemotherapy. *Acta Orthop Scand.* (1998) 69:230–6. doi: 10.3109/17453679809000921
34. Arshi A, Sharim J, Park DY, Park HY, Yazdanshenas H, Bernthal NM, et al. Prognostic determinants and treatment outcomes analysis of osteosarcoma and Ewing sarcoma of the spine. *Spine J.* (2017) 17:645–55. doi: 10.1016/j.spinee.2016.11.002
35. Lee J, Hoang BH, Ziogas A, Zell JA. Analysis of prognostic factors in Ewing sarcoma using a population-based cancer registry. *Cancer.* (2010) 116:1964–73. doi: 10.1002/cncr.24937
36. Bielack SS, Kempf-Bielack B, Delling G, Exner GU, Flege S, Helmke K, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol.* (2002) 20:776–90. doi: 10.1200/JCO.2002.20.3.776

37. Berner K, Bruland ØS. Prognostic impact of proximal versus distal localization in extremity long bone osteosarcomas. *Anticancer Res.* (2019) 39:2459–66. doi: 10.21873/anticanres.13365
38. Wan ZH, Huang ZH, Chen LB. Survival outcome among patients with Ewing's sarcoma of bones and joints: a population-based cohort study. *São Paulo Med J.* (2017) 136:116–22. doi: 10.1590/1516-3180.2017.0236230917
39. Oberlin O, Deley MC, Bui BN, Gentet JC, Philip T, Terrier P, et al. Prognostic factors in localized Ewing's tumours and peripheral neuroectodermal tumours: the third study of the French Society of Paediatric Oncology (EW88 study). *Br J Cancer.* (2001) 85:1646–54. doi: 10.1054/bjoc.2001.2150
40. Bieling P, Rehan N, Winkler P, Helmke K, Maas R, Fuchs N, et al. Tumor size and prognosis in aggressively treated osteosarcoma. *J Clin Oncol.* (1996) 14:848–58. doi: 10.1200/JCO.1996.14.3.848
41. Biazzo A, De Paolis M. Multidisciplinary approach to Osteosarcoma. *Acta Orthop Belg.* (2016) 82:690–8.
42. DuBois SG, Krailo MD, Gebhardt MC, Donaldson SS, Marcus KJ, Dormans J, et al. Comparative evaluation of local control strategies in localized Ewing sarcoma of bone: a report from the Children's Oncology Group. *Cancer.* (2015) 121:467–75. doi: 10.1002/cncr.29065
43. PDQ Pediatric Treatment Editorial Board. Ewing Sarcoma Treatment (PDQ®): Patient Version. In: *PDQ Cancer Information Summaries*. Bethesda (MD): National Cancer Institute (US); December 22, 2020.
44. Ellis MA, Gerry DR, Neskey DM, Lentsch EJ. Ewing sarcoma of the head and neck. *Ann Otol Rhinol Laryngol.* (2017) 126:179–84. doi: 10.1177/0003489416681322
45. Duchman KR, Gao Y, Miller BJ. Prognostic factors for survival in patients with Ewing's sarcoma using the surveillance, epidemiology, and end results (SEER) program database. *Cancer Epidemiol.* (2015) 39:189–95. doi: 10.1016/j.canep.2014.12.012
46. Karski EE, McIlvaine E, Segal MR, Krailo M, Grier HE, Granowetter L, et al. Identification of discrete prognostic groups in ewing sarcoma. *Pediatr Blood Cancer.* (2016) 63:47–53. doi: 10.1002/pbc.25709

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Zhanghuang, Tan, Mi, Liu, Jin, Li, Zhang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fpubh.2022.834032

Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network

Ahila A[1]*, *Poongodi M*[2]*, *Mounir Hamdi*[2], *Sami Bourouis*[3], *Kulhanek Rastislav*[4] and *Faizaan Mohmed*[5]

[1] Department of Electronics and Communication Engineering, Sethu Institute of Technology, Kariapatti, India, [2] College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar, [3] Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, [4] Information Systems Department, Faculty of Management, Comenius University in Bratislava, Bratislava, Slovakia, [5] School of Creative Tech, University of Bolton, Bolton, United Kingdom

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Rathinaraja Jeyaraj,
Kyungpook National University, South Korea
Praveen Kumar,
VIT University, India
Mukund Janardhanan,
University of Leicester, United Kingdom

***Correspondence:**
Ahila A
akhilaamarnath27@gmail.com
Poongodi M
dr.m.poongodi@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 12 December 2021
Accepted: 05 January 2022
Published: 07 February 2022

Citation:
A A, M P, Hamdi M, Bourouis S, Rastislav K and Mohmed F (2022) Evaluation of Neuro Images for the Diagnosis of Alzheimer's Disease Using Deep Learning Neural Network. Front. Public Health 10:834032. doi: 10.3389/fpubh.2022.834032

Alzheimer's Disease (AD) is a progressive, neurodegenerative brain disease and is an incurable ailment. No drug exists for AD, but its progression can be delayed if the disorder is identified at its initial stage. Therefore, an early analysis of AD is of fundamental importance for patient care and efficient treatment. Neuroimaging techniques aim to assist the physician in the diagnosis of brain disorders by using images. Positron emission tomography (PET) is a kind of neuroimaging technique employed to create 3D images of the brain. Due to many PET images, researchers attempted to develop computer-aided diagnosis (CAD) to differentiate normal control from AD. Most of the earlier methods used image processing techniques for preprocessing and attributes extraction and then developed a model or classifier to classify the brain images. As a result, the retrieved features had a significant impact on the recognition rate of previous techniques. A novel and enhanced CAD system based on a convolutional neural network (CNN) is formulated to address this issue, capable of discriminating normal control from Alzheimer's disease patients. The proposed approach is evaluated using the 18FDG-PET images of 855 patients, including 635 normal control and 220 Alzheimer's disease patients from the ADNI database. The result showed that the proposed CAD system yields an accuracy of 96%, a sensitivity of 96%, and a specificity of 94%, leading to splendid performance when related to the methods already in use that are specified in the literature.

Keywords: Alzheimer's disease, accuracy, convolutional neural network, deep learning, feature extraction, image analysis, image classification and positron emission tomography

1. INTRODUCTION

Alzheimer's disease (AD) is a type of brain disease which regularly affects people over 65 years old. It is a progressive and neurodegenerative disorder, meaning that, it becomes worse with time. According to an Alzheimer's Association report, around 55 million people are living with AD worldwide and it is envisioned that the number of AD patients will reach 152 million by 2,050. The number of people with AD is progressively increasing worldwide. The prime objective of this paper is to develop a robust classification system for AD diagnosis using a convolutional neural network (CNN). In this approach, the 3D image classification problem is completely

solved by converting them into 2D images. The proposed system uses whole brain images for AD diagnosis instead of region of interest. Thus, there is no need for a segmentation algorithm. Furthermore, the proposed system uses CNN for classification. CNN learns the features from images. Pre-processing like noise removal, enhancement, and feature extraction are not needed. AD starts with impairment of memory functions followed by cognitive functions with behavioral impairments (1). Some common symptoms of AD are: memory loss, a lack of initiative, difficulty in expressing thoughts and recognizing people or relations, personality changes, and poor judgement, etc. AD is an incurable brain disorder. Currently, no effective treatment has yet been fully discovered, but it is possible to slow down the progression of AD if the disorder is identified at its initial stage. Therefore, it is of primary significance to identify AD at an early stage for patient care, as well as effective treatment. Furthermore, if the condition is diagnosed early on, some of the changes induced by AD can be reversed, allowing patients to keep their everyday lives. However, diagnosis of AD still remains a highly complicated task, especially at an early stage while the disorder offers greater opportunities to be treated (2).

Over the past few years, several neuro imaging modalities which include magnetic resonance imaging (MRI), positron emission tomography (PET) (3), and single-photon emission computerized tomography (SPECT) have been confirmed to be very powerful within the diagnosis of AD. PET is a non-invasive neuro imaging technique that uses an imaging agent like 18FDG-PET to monitor the brain's glucose intake. Brain activity is related to glucose consumption. During PET scanning, a position emitting radionuclide tracer with 18FDG is delivered in the body. The glucose uptake shows tissue enzymatic reactions when the quantities of these traces are scanned with a camera. In seriously damaged AD patients, particular brain areas showed reduced glucose intake, together with bilateral areas inside the temporal and partial lobes, posterior cingulate, frontal lobes, and entire brain (4). Research studies proved that FDG-PET is a good candidate for AD diagnosis and can be used to assist the physicians or an expert to analyze and diagnosis AD in early stages. Consequently, this paper proposes an automated computer-aided design (CAD) system for the diagnosis of AD using 18FDG-PET images. The foremost awareness of this research work is to distinguish AD patients from normal control (NC) employing CNN (5).

Classification of medical images via visual examination by an expert or a physician can be subjective and prone to errors. As a result, researchers attempted to create an automatic CAD system that could distinguish AD from NC participants using picture data attributes. Over the past few years, numerous machine learning models have evolved and are being used for examining neuro reading that allows researchers to capture the structure or functional changes related to AD including support vector machine (SVM) (6) classification in SPECT and learning vector quantization-SVM (LVQ-SVM) classification in structural MRI (7).

Wysoczánski et al. (1) investigated the strength of random forest (RF) for diagnosing AD. The authors used partial least square (PLS) for extracting features from the image. This method is tested on SPECT images downloaded from the ADNI dataset. A new CAD system for differentiating AD from NC based on PET image features was developed by Cabral and, Silveria (4). Garali et al. (8) proposed a classification system using favorite class ensembles for differentiating AD and MCI from NC using PET images. Each classifier in the ensemble uses a different set of brain voxels. Two base classifiers SVM and random forest are employed to create the ensemble classifier. Results showed that this ensemble classifier outperforms the corresponding single classifier. However, higher computational cost is the drawback of this system. Markiewicz et al. (9) presented a region-based method to classify AD from NC using 16 anatomical regions of interest (ROIs), and the first four times, as well as their entropy, are computed and used as feature vectors. The capacity of ROIs to discriminate PET images is then ranked using ROC curves. Finally, 21 capabilities are selected and fed as an input to both SVM and RF classifiers. Results proved that this method achieved high classification accuracy using 166 anatomical ROIs when compared to other methods. Poongodi and Bose (10) proposed a deep learning model for AD diagnosis from MRI images. A deep neural network is designed with a convolution layer, normalization layer, and pooling layer. The effectiveness of this method is validated on the OASIS dataset. Results showed improved performance of this particular method when compared to other methods. A boosting classifier to classify PET images was developed by Poongodi and Bose (11). It is the combination of simple classifiers, which performs segmentation, feature extraction, and feature selection in order to make the input image fit for the classification task. This classifier was validated on the ADNI dataset and achieved an accuracy of 90.97%.

2. DISTINCT FEATURES OF THE PROPOSED METHOD

In the literature, researchers have developed an automated CAD system using various soft computing models (12). Most of the researchers have used either full brain images or set features for diagnosing AD. Each method has its own merits and demerits. However, none of the methods offer a consistent outcome. Keeping this in mind, this research work intends to design and render a CAD system for discriminating AD from NC with the help of CNN. In the proposed CAD system, the first 15 slices and last 15 slices are discarded to remove unwanted regions. Dimension-reduced 18FDG-PET images are fed as input into the deep neural network. CNN is used for classifying FDG-PET snapshots into NC and AD. The efficacy of CNN is evaluated by measuring accuracy, sensitivity, and specificity, and is compared with the existing methods (13–16).

For solving the shortcomings of the methods elaborated in the literature, this paper presents a novel computer diagnosis system using a deep convolutional neural network. The conventional neural network has the potential of recognizing images with minimal processing. The novel features of the proposed system are: 3D images are transformed into 2D images so as to reduce the computational time, 2D images are further grouped at specific intervals for reducing the dimension, image segmentation is not

required in this proposed approach, and no image processing techniques are needed to extract features from the images.

From the existing literature survey, it could be seen that, in existing computer-aided diagnosis systems, in those intended to classify the brain images (5), it was observed that most of the systems employ image processing methods to classify the data, therefore, a machine learning model was designed to obtain the features. Also, few methods use only the information from regions of interest (17). However, these methods require increased computational time. In addition, the classification accuracy depends mainly upon the selected features. Researchers developed hybrid models by combining a deep learning neural network and machine learning model to increase the classification accuracy, but these models required increased memory usage and increased time (18).

To overcome these issues, the proposed CAD system employs the consideration of entire brain images instead of regions of interest. The 3D picture type problem is completely solved with the aid of converting them into 2D images. Venugopalan et al. (19) proposed a set of rules that might help in robot surgical treatment. To run on such soft tissues, software program-driven techniques and algorithms have to be extra particular in choosing the highest quality direction for attaining the procedural region. Statistical analysis has determined whether the proposed approach might be outperforming under the favorable learning rate, discount factor, and the exploration factor (20). The network is built with multiple layers to learn features through a training process, which eliminates the need for extracting the features, resulting in higher prediction performance when compared to other approaches. Reddy et al. (21) focused on classifying tomato disease with a machine learning model used to predict agricultural disaster. The features were extracted from the dataset using the hybrid-predominant element evaluation-whale optimization algorithm and the extracted features were fed into a deep neural network for classification of tomato diseases. Srinivasu et al. (22) focused on improving quality, a preprocessing technique was implemented in a multimodal stroke dataset from the Kaggle repository. To achieve homogeneity, a label encoder technique was used and dataset missing values were replaced with attribute means. Resampling techniques were used to balance the dataset and to obtain accurate results. The following is how the rest of the paper is organized: The demographic details of the dataset utilized in this work are presented in section 2. Section 3 formulates the unique features of the proposed CAD device for AD prognosis. Section 4 presents the unique evaluation of numerical results and evaluation of the results. Finally, section 5 concludes the paper with few suggestions toward future research in closer vicinity of this study.

3. DATASET DETAIL

The neuroimaging dataset that was used in this study is presented in this section.

The visual dataset used in this study was acquired from ADNI, which is publicly available (http://adni.loni.usc.edu/methods/pet-analysis-methods). The ADNI database was released in 2003 by the NIA, NIBIB, and FDA. The ADNI was designed for the diagnosis of AD at the earliest stage and tracking of progression of AD with several biomarkers such as MRI, PET, and SPECT. ADNI helps researchers to gather, collect, and use the image data including PET to measure and track the progression of AD.

TABLE 1 | Demographic details of 18FDG-PET image dataset.

Factors	Normal control	Alzheimer's disease
Number of subjects	635	220
Age	70	75
Gender (M/F)	35/15	30/20
MMSE	24–30	20–26
CDR	0	1.5 or 1

18FDG-PET images of 855 patients and their final diagnosis were collected from ADNI. The detailed procedure for 18FDG-PET scanning can be found on the website (http://adni.loni.usc.edu/methods/pet-analysis-methods). The total data have 855 samples including 635 NC and 220 AD subjects. The dataset is split into two categories: training data and validation data. Around 90% of the data, referred to as training data, was used to construct the model, while the remaining 10%, referred to as validation data, was used to validate the model. Description of 18FDG-PET images investigated in this work is provided in **Table 1**. Only subjects with CDR of 0.5 or 1 for AD and 0 for NC were chosen. Classification was based on MMSE (Mini Mental State Examination) scores of 24 to 30 for NC and 22 to 26 for AD. The resulting volumes of images have been represented by means of a matrix of size 256x256x96 yielding a total of 6291456 voxels. The values are denoted as mean plus or minus std deviation.

4. PROPOSED CAD SYSTEM

4.1. In-linestyle

The capture of 18FDG-PET images in ADNI followed a conventional methodology. The pictures were co-registered, averaged, aligned, normalized, and smoothed to a similar resolution of 8 mm FWHM 30–60 min after injection. Each image was examined for artifacts and, if needed, its orientation was changed.

Subsequently, the FDG-PET images have been normalized through an affine model with 12 parameters using SPM12 software, normalized in intensity, and converted to a uniform isotropic resolution of 8 mm FWHM. All the images were reduced to a size of 160 x 160 x 96 for analysis. **Figure 1** shows the sample of co-registration using SPM12.

4.2. Classification

Numerous methods have been proposed for AD diagnosis using various approaches like RF, SVM, and ANN. Each method has its own characteristics. In recent years, deep learning neural networks were used for medical photo evaluation and computer vision problems due to its ability to capture the features from 2D images. Additionally, CNN has the potential to recognize

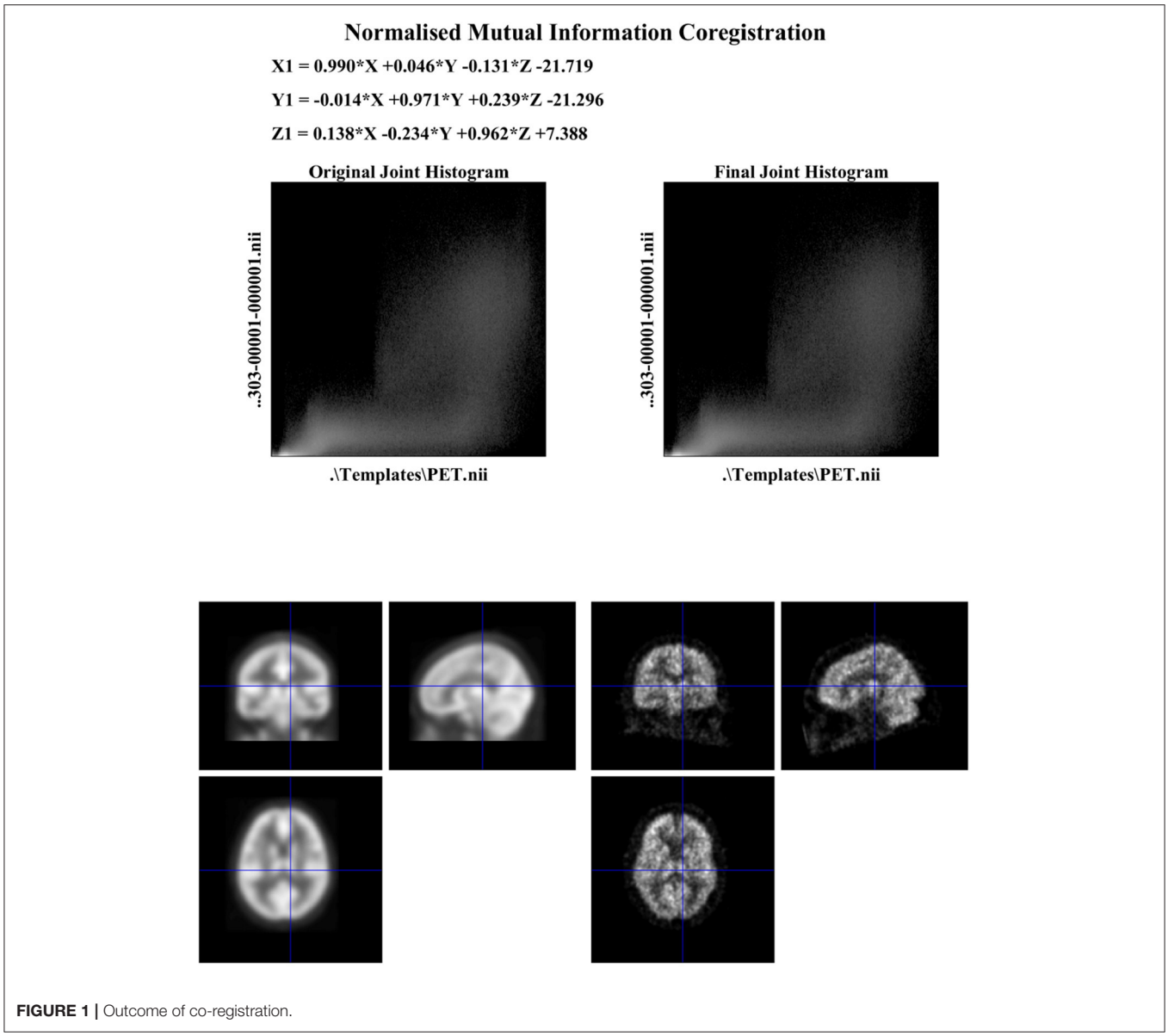


FIGURE 1 | Outcome of co-registration.

visual patterns with minimal preprocessing. CNN is robust to spatial and geometric transformations. In this paper, a 2D CNN is built to differentiate AD from NC. **Figure 2** depicts the suggested CAD system's framework. Each 3D FDG-PET image is transformed into multiple 2D images along the coronal (axial) direction, as shown in **Figure 2**. The initial 15 and closing 15 slices are discarded to do away with skull and other undesirable regions. Subsequently, the 2D slices are arranged into groups at particular intervals with some overlaps and averaged to reduce the dimension and computational overhead. CNN is developed with many layers and trained to capture useful features. The features generated by convolution layers are implemented to absolutely connected layers followed by means of a smooth max to make the final decision.

The convolution layer, pooling or subsampling layer, fully connected layer, and soft max classification layer are the layers that make up a CNN. The designed 2D CNN is composed of an input layer, three convolutional layers, two subsampling or pooling layers, a drop out layer, a fully connected layer, and a smooth max layer as shown in **Figure 3**. Parameters employed for constructing the CNN are tabulated in **Table 2**. The hyper parameters are tuned by experimentation. After many trails, the values are fixed. The designed CNN has an input layer size of 160 x 160. The sizes of receptive fields or filters for convolutional layers are set to 3 x 3 and kernel numbers are 8, 16, and 32 for C1, C2, and C3, respectively. Convolutional layers are used for extracting features from the image. The first C1 generates low level features with a size of 158 x 158. The pooling layer is used to reduce the dimension of the features. Max pooling function is adopted. Subsequent to P1, these features are reduced into 79 x 79. Then, C2 generates mid-level features with a size of 39 x 39 and these features are down-sampled into 20 x 20. Finally, C3

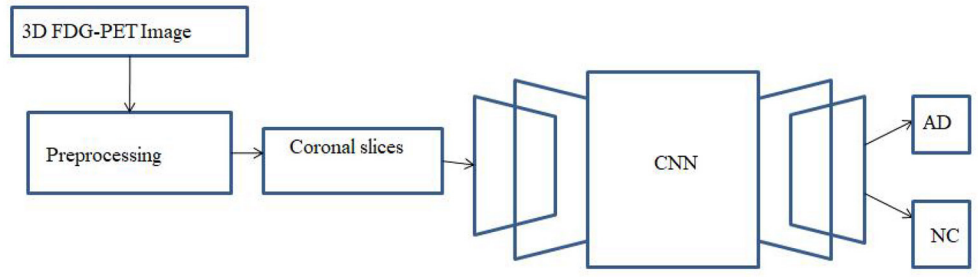


FIGURE 2 | Framework of the proposed CAD system.

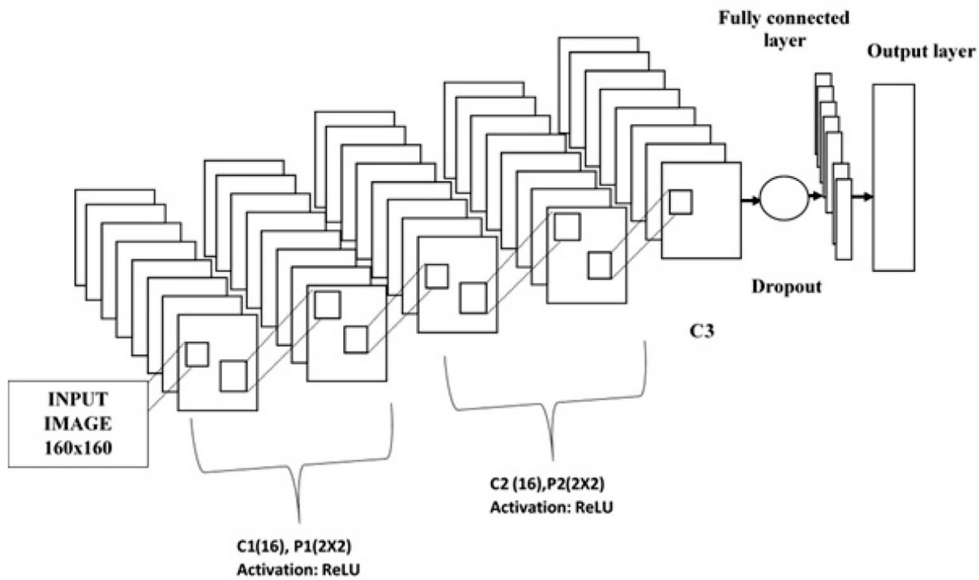


FIGURE 3 | Convolutional neural network.

generates high-level features with a size of 10 x 10. These features are concatenated and taken as the feature vector. The obtained feature vector is fed to FC followed by the soft max layer. In this work, the rectified linear unit (ReLU) activation function is adopted to introduce non-linearity. SGDM is used to train the network by minimizing the cross entropy loss, and maximum iteration is set to 500. The drop out layer is utilized to improve the generalization ability and to prevent the overfitting problem.

5. NUMERICAL RESULTS AND DISCUSSION

This section depicts the numerical results of the proposed CAD system for AD diagnosis from FDG-PET images. It also goes through the simulation environment and performance measures that were used to compare the proposed system to other existing approaches.

TABLE 2 | CNN parameters employed.

Sl.No.	Layer	Kernel size/stride	No. of kernels	Size
1	Input image			160 x 160 x 1
2	Convolution layer 1 (C1)	3 x 3/1	16	158 x 158 x 16
3	Pooling layer 1 (P1)	2 x 2/2		79 x 79 x 16
4	Convolution layer 2 (C2)	3 x 3/1	32	39 x 39 x 32
5	Pooling layer 2 (P2)	2 x 2/2		20 x 20 x 32
6	Convolution layer 3 (C3)	3 x 3/2	64	10 x 10 x 64
7	Flatten			6,400
8	Fully connected layer	512		512
9	Soft max	2		2

5.1. Simulation Platform

The main objective of simulation platform is to provide better results for visualization and interpretation. The developed CAD

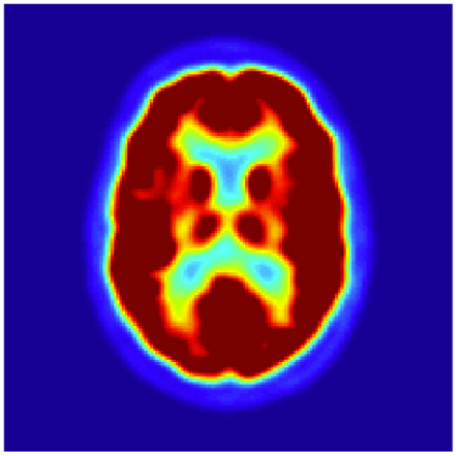


FIGURE 4 | Sample of preprocessed NC images.

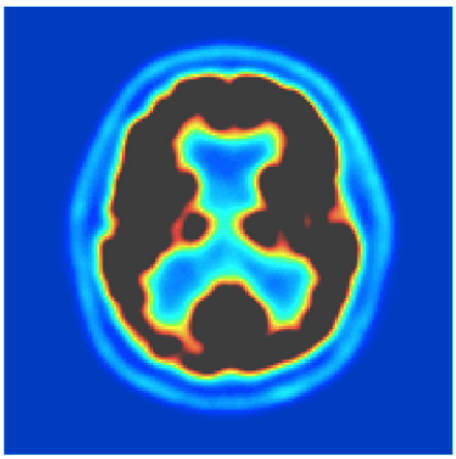


FIGURE 5 | Sample of preprocessed AD images.

system is implemented on the MATLAB 2018a platform using the deep learning toolbox and executed in an Intel core i5 processor with 2.5 GHz speed and 12 GB of RAM.

5.2. Performance Indicators

To estimate the efficacy of the proposed CAD system, this paper evaluates some widely used performance indicators, such as confusion matrix, region of curve (ROC), classification accuracy, sensitivity, and specificity. Accuracy calculates the proportion between true positive (correctly classified data) and total data, and shall be mathematically represented as given in Equation (1).

Specificity is the percentage of normal control patients who are identified as not having AD, and can be expressed as mentioned in Equation (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2)$$

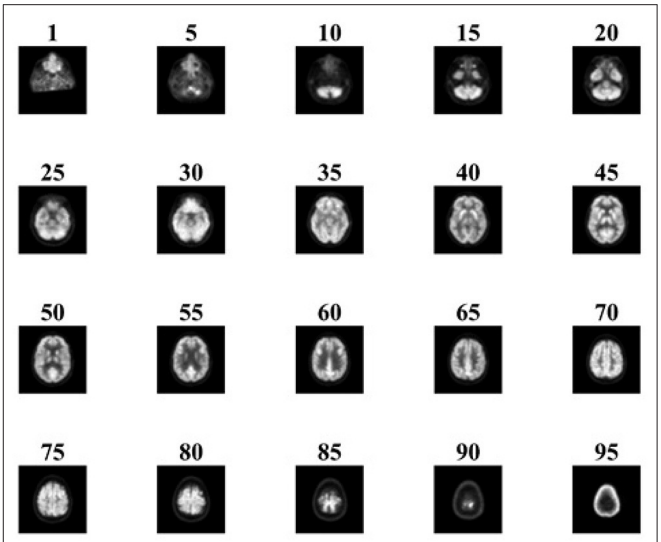


FIGURE 6 | Sample slices of NC patient.

Sensitivity computes the percentage of AD patients who are diagnosed and as such sensitivity is calculated as expressed in Equation (3).

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

True positive is represented by TP, true negative is represented by TN, false negative is represented by FN, and false positive is represented by FP.

5.3. Experimental Treatment

In this work, an automated CAD system for the diagnosis of brain disorders is developed and implemented accordingly. The proposed system was employed to differentiate AD patients from NC. The FDG-PET image dataset was utilized in the simulation and was gathered from the ADNI database. The FDG-PET images were acquired from 855 patients including 635 NC and 220 AD patients.

The collected images are preprocessed using SPM12. The samples of preprocessed images of NC and AD patients are given in **Figures 4**, **5**, respectively. It is obvious that hypometabolism is visible in AD patients. Moreover, the first 15 and last 15 slices have redundant information. Slices of these regions of AD and NC patients are demonstrated in **Figures 6**, **7**, respectively. These slices are discarded to remove unwanted regions and make the image fit for further processing. Preprocessed and dimension-reduced images are fed as input into the designed system. CNNs are a kind of deep learning neural network that have proven to be very effective in image classification recognition problems. The CNN learns features from the images acquired using its own convolutional layer. Additionally, CNNs can recognize patterns with extreme variability and some geometric transformations such as scaling, rotation, translation, and noise.

Figure 8 shows the training plot of the CNN. The outcome of the proposed CAD system is demonstrated in the form of a

confusion matrix as detailed in **Figure 9**. The total number of images is 855 including 635 NC and 220 AD patients.

To demonstrate the superior performance, the developed system was repeated five times and the mean values were reported. Efficacy of the developed system is evaluated by computing evaluation parameters which are given in section 2. The outcome of the proposed CAD system for the testing sample is demonstrated in the form of a confusion matrix as detailed in **Figure 9**.

In **Figure 9**, 42 images are correctly diagnosed as NC patients and 41 images are correctly classified as AD patients. A total of 2 images out 43 images are mistakenly identified as AD (2.3%). Similarly, one image is misdiagnosed as NC (1.2%). For NC patients, out of 43 patients, 48.8% are correctly diagnosed as NC and 2.3% are wrongly diagnosed. For AD patients, out of 43 patients, 47.7% are correctly classified as AD and 1.2% are wrongly classified. Overall, 96.5% of the photos are correctly

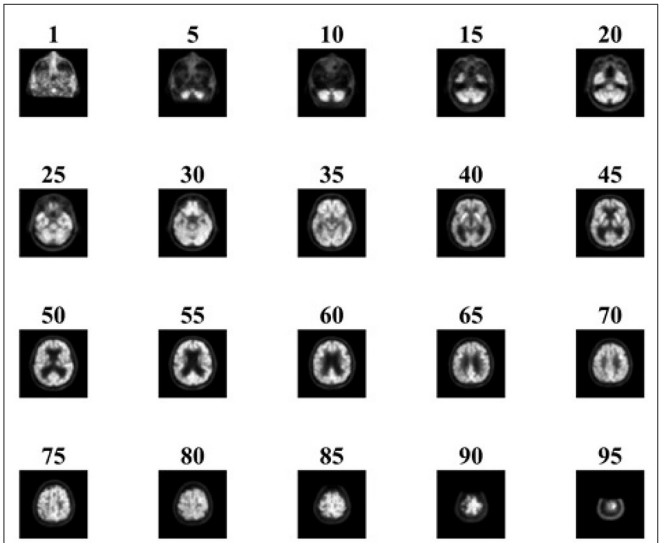


FIGURE 7 | Sample slices of AD patient.

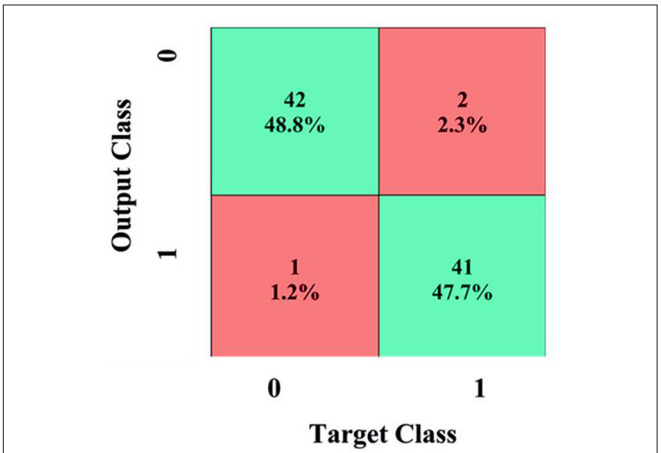


FIGURE 9 | Confusion matrix.

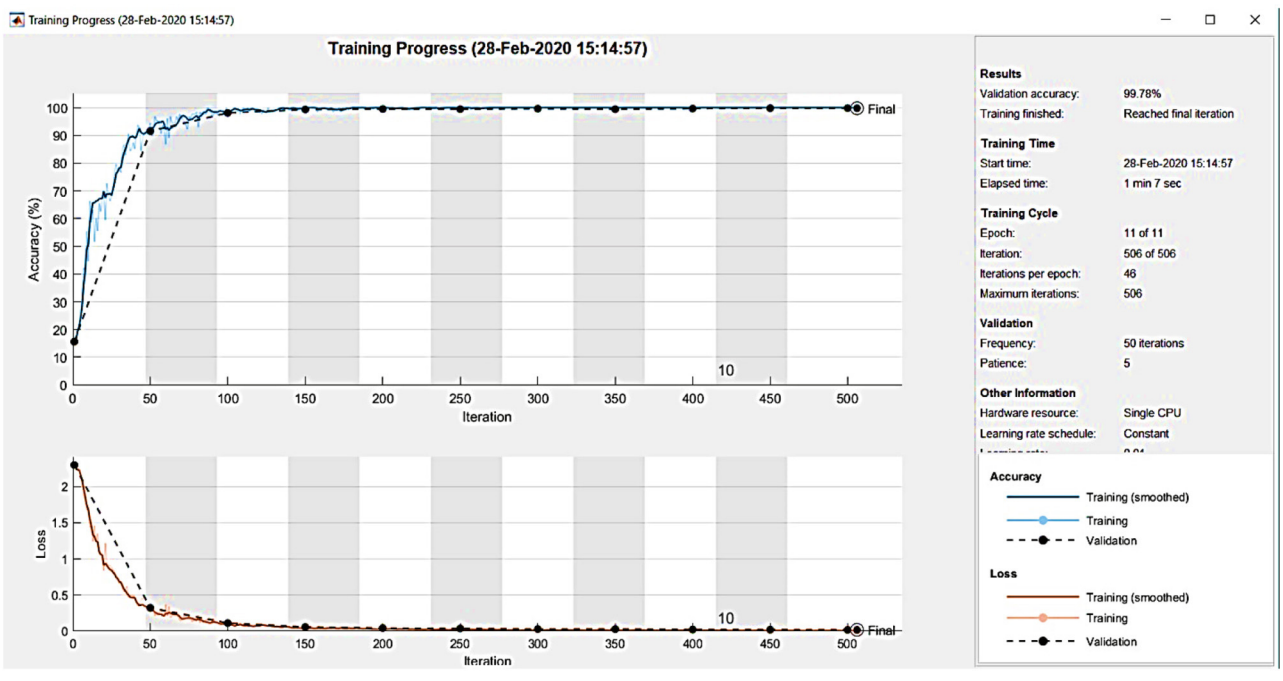


FIGURE 8 | CNN training process.

classified, whereas 4.5% are incorrectly categorized. ROC is a mathematical tool used to assess the separation ability of the classification system. The true positive rate is compared against the false positive rate in a ROC chart. It is investigated by adjusting the threshold value. **Figure 10** shows the ROC of the proposed system. The area under the curve (AUC) of the proposed system is 0.95.

5.4. Effect Analysis

In terms of accuracy, sensitivity, and specificity, this section compares the efficacy of the proposed CAD system to that of existing approaches. In particular, the experimental consequences reported in state-of-the-art methods such as Cabral and Silveria (4), Liu et al. (23), Islam and Zhang (24), Silveira and Marques (25), and Poongodi and Bose (26) are compared with the proposed system, as given in **Table 3**.

Cabral and Silveria (4) proposed a method to differentiate NC from AD using PET images. Input images were preprocessed using SPM. Feature vectors were extracted with PCA. SVM classifier discriminated NC from AD based on the feature vectors. Regional analysis-based classification system was proposed by Liu et al. (23). In this approach, the PET image was preprocessed using SPM. SVM classifier was used to carry out the classification process. Islam and Zhang (24) investigated the strength of CNN for diagnosing AD. They converted the 3D MRI images into 2D images. To increase the image's quality, a few preprocessing techniques, such as noise reduction, edge recognition, and segmentation were used. The preprocessed image was fed as an input to the CNN. The deep learning neural network provided a better result compared to the existing methods.

Silveira and Marques (25) presented a classification system based on a CNN and extreme machine learning classifier. CNN was utilized to extract characteristics from MRI images in this system. Then, with the help of free surfer, the features were mined. Extreme machine learning classifier performed the classification task. The drawback of the system is high computational overhead. A hybrid method of CNN and RNN for AD diagnosis was presented by Poongodi and Bose (26). In this method, 3D PET images were transformed into 2D slices along axial, coronal, and sagittal directions to reduce the computation time. At regular intervals, the transformed slices were divided into a number of groups. The resultant groups were fed as input to the CNN. To extract inter slice features, RNN was used. Prediction score was calculated using the weighted averaging method. However, this method needs more memory and time. **Figure 11** shows a graphic comparison of the suggested CAD system's classification with the current approaches.

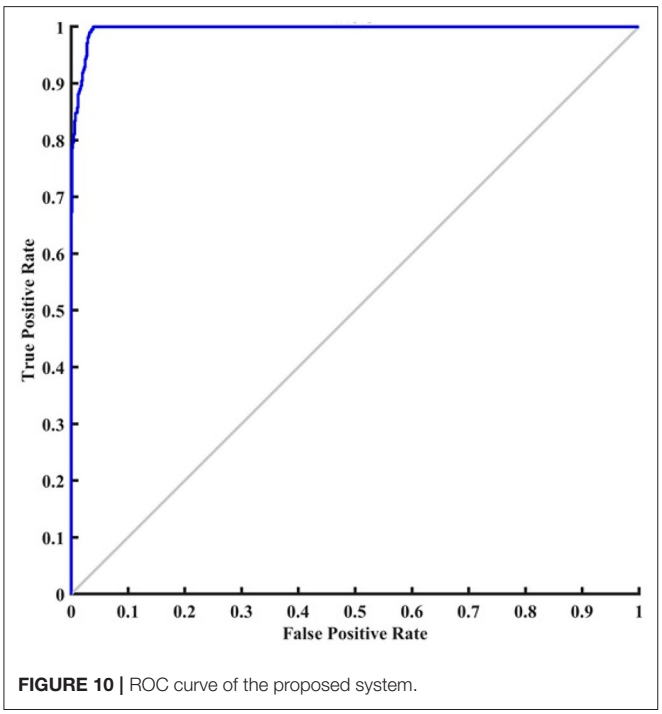


FIGURE 10 | ROC curve of the proposed system.

6. CONCLUSION AND FUTURE WORKS

In this paper, a CAD device for discriminating AD from NC patients based on features from 18FDG-PET images was proposed and investigated properly. The proposed CAD system was designed by using a convolutional neural network. The FDG-PET images were decomposed into several 2D slices for extracting the features. Subsequently, the slices were grouped at some intervals without overlaps. The proposed CAD machine was confirmed on the ADNI dataset. The obtained simulation results showed that the proposed system provided exceptional performance when compared to the existing methods. Simulation results clearly demonstrated that the proposed CNN possessed higher potential to differentiate

TABLE 3 | Performance comparison of the proposed system with existing systems.

Author	Modality	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
Cabral and Silveria (4)	PET	PCA-SVM	74.18	92.75	15.91
Liu et al. (23)	FDG-PET	SVM	90	82.7	80.4
Islam and Zhang (24)	MRI	CNN	95.3	84.4	71.4
Silveira and Marques (25)	MRI	DNN	86.1	84	74.8
Poongodi and Bose (26)	FDG-PET	CNN-RNN	95.3	91.4	91
Proposed	FDG-PET	CNN	96.8	94	96

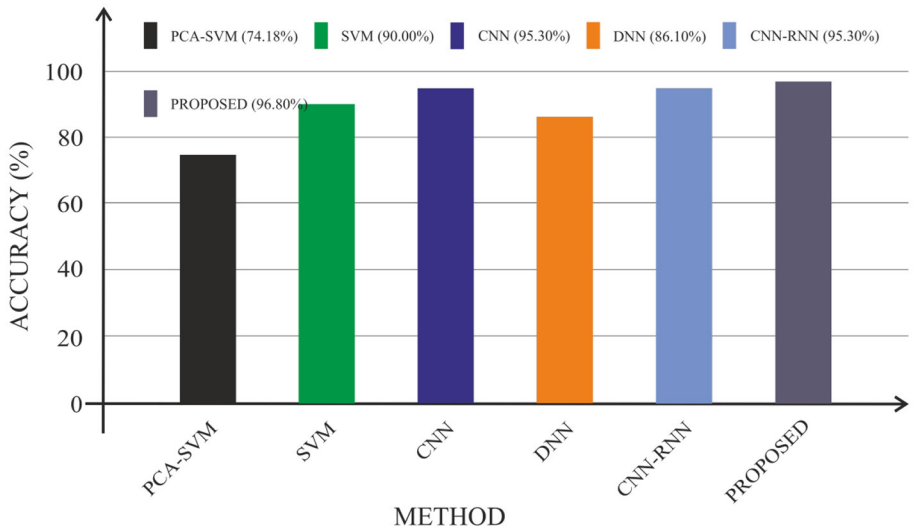


FIGURE 11 | Performance comparisons in terms of accuracy.

AD from NC with better classification accuracy and robustness. Future study will focus on improving the classification accuracy by extending the suggested CAD system to incorporate data from other sources. The overall performance of the suggested CAD system will be evaluated using a large number of samples. In addition, the potential of using different deep learning neural networks will be explored in a wider scenario.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AA: conceptualization, original draft writing, and implementation. PM and MH: funding acquisition. SB: supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Taif University Researchers Supporting Project (number TURSP-2020/26), Taif University, Taif, Saudi Arabia.

REFERENCES

1. Wysoczański D, Mroczka J, Polak AG. Performance analysis of regularization algorithms used for image reconstruction in computed tomography. *Bull Polish Acad Sci Techn Sci.* (2013) 61:467–74. doi: 10.2478/bpasts-2013-0046
2. Poongodi M, Sharma A, Vijayakumar V, Bhardwaj V, Sharma AP, Iqbal R, et al. Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Comput Elect Eng.* (2020) 81:106527. doi: 10.1016/j.compeleceng.2019.106527
3. Grochowski M, Kwasigroch A, Mikołajczyk A. Selected technical issues of deep neural networks for image classification purposes. *Bull Polish Acad Sci Tech Sci.* (2019) 61:363–76. doi: 10.24425/bpas.2019.128485
4. Cabral C, Silveria M. Classification of Alzheimer's disease from FDG-PET images using Favorite class ensemble. In: *Proc. 35th International Conference of the IEEE EMBS Engineering in Medicine and Biology Society.* Osaka (2013). p. 2477–80.
5. Association A. 2019 Alzheimer's disease facts and figures. *Alzheimer's Dementia.* (2019) 15:321–87. doi: 10.1002/alz.12068
6. Stasiak B, Tarasiuk P, Michalska I, Tomczyk A. Application of convolutional neural networks with anatomical knowledge for brain MRI analysis in MS patients. *Bull Polish Acad Sci Tech Sci.* (2018) 66:857–68. doi: 10.24425/bpas.2018.125933
7. Illán IA, Górriz IM, Ramírez J, Gonzalez DS. 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis". *Inf Sci.* (2011) 181:903–16. doi: 10.3389/fnagi.2018.00158
8. Garali I, Adel M, Bourennane S, Guedj E. Region based brain selection and classification on PET images for Alzheimer's disease computer aided diagnosis. In: *Proc IEEE International Conf on Image Processing.* Quebec City, QC (2015). p. 1473–7.
9. Markiewicz T, Swiderska-Chadaj Z, Gallego J, Bueno G, Grala B, Lorent M. Deep learning for damaged tissue detection and segmentation in Ki-67 brain tumor specimens based on the U-net model. *Bull Polish Acad Sci Tech Sci.* (2018) 66:849–56. doi: 10.24425/BPAS.2018.125932
10. Poongodi M, Bose S. A firegroup mechanism to provide intrusion detection and prevention system against DDoS attack in collaborative clustered networks". *Int J Inf Security Privacy.* (2014) 8:1–18. doi: 10.4018/IJISP.2014040101
11. Poongodi M, Bose S. Stochastic model: reCAPTCHA controller based co-variance matrix analysis on frequency distribution using trust evaluation and re-eval by Aumann agreement theorem against DDoS attack in MANET." *Clust Comput.* (2015) 18:1549–59. doi: 10.1007/s10586-015-0496-y
12. Gray KR, Wolz R, Keihaninejad S, Heckemann RA, Aljabar P, Hammers A. Regional analysis of FDG-PET for use in the classification of Alzheimer's disease. In: *Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro.* Chicago, IL (2011). p. 1082–5.

13. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Publ Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
14. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-Technological factors affecting user's adoption of eHealth functionalities: A case study of China and Ukraine eHealth systems. *IEEE Access.* (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
15. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-sanitization: a semantic privacy-preserving framework for unstructured medical datasets. *Comput Commun.* (2020) 161:160–71. doi: 10.1016/j.comcom.2020.07.032
16. Patel H, Singh Rajput D, Thippa Reddy G, Iwendi C, Kashif Bashir A, Jo O. A review on classification of imbalanced data for wireless sensor networks. *Int J Distrib Sensor Netw.* (2020) 16:1550147720916404. doi: 10.1177/1550147720916404
17. Lei B, Liang E, Yang M, Yang P, Zhou F, Tan EL, et al. Predicting clinical scores for Alzheimer's disease based on joint and deep learning. *Exp Syst Appl.* (2022) 187:115966. doi: 10.1016/j.eswa.2021.115966
18. Lin E, Lin CH, Lane HY. Deep learning with neuroimaging and genomics in Alzheimer's disease. *Int J Mol Sci.* (2021) 22:7911. doi: 10.3390/ijms22157911
19. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep.* (2021) 11:1–13. doi: 10.1038/s41598-020-74399-w
20. He Y, Wu J, Zhou L, Chen Y, Li F, Qian H. Quantification of Cognitive Function in Alzheimer's Disease Based on Deep Learning. *Front Neurosci.* (2021) 15:178. doi: 10.3389/fnins.2021.651920
21. Reddy T, Bhattacharya S, Maddikunta PKR, Hakak S, Khan WZ, Bashir AK, et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimedia Tools Appl.* (2020) 5:1–25. doi: 10.1007/s11042-020-09988-y
22. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors. *J Real Time Image Process.* (2021) 18:1773–85. doi: 10.1007/s11554-021-01122-x
23. Liu M, Cheng D, Yan W. Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images. *Neuro Inf.* (2018) 12:35. doi: 10.3389/fninf.2018.00035
24. Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Inf Springer.* (2018) 5:2. doi: 10.1186/s40708-018-0080-3
25. Silveira M, Marques J. Boosting Alzheimer disease diagnosis using PET images. In: *Proc. 20th IEEE International Conference on Pattern Recognition (ICPR).* Istanbul (2015). p. 2556–9.
26. Poongodi M, Bose S. Detection and Prevention system towards the truth of convergence on decision using Aumann agreement theorem. *Proc Comput Sci.* (2014) 50:244–51. doi: 10.1016/j.procs.2015.04.053

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 A, M, Hamdi, Bourouis, Rastislav and Mohmed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fpubh.2021.804471

Energy Expenditure Estimation of Tabata by Combining Acceleration and Heart Rate

*Yiping Yan and Qingguo Chen**

Sport Laboratory, College of Physical Education, Sichuan Normal University, Sichuan, China

Tabata training plays an important role in health promotion. Effective monitoring of exercise energy expenditure is an important basis for exercisers to adjust their physical activities to achieve exercise goals. The input of acceleration combined with heart rate data and the application of machine learning algorithm are expected to improve the accuracy of EE prediction. This study is based on acceleration and heart rate to build linear regression and back propagate neural network prediction model of Tabata energy expenditure, and compare the accuracy of the two models. Participants ($n = 45$; Mean age: 21.04 ± 2.39 years) were randomly assigned to the modeling and validation data set in a 3:1 ratio. Each participant simultaneously wore four accelerometers (dominant hand, non-dominant hand, right hip, right ankle), a heart rate band and a metabolic measurement system to complete Tabata exercise test. After obtaining the test data, the correlation of the variables is calculated and passed to linear regression and back propagate neural network algorithms to predict energy expenditure during exercise and interval period. The validation group was entered into the model to obtain the predicted value and the prediction effect was tested. Bland-Alterman test showed two models fell within the consistency interval. The mean absolute percentage error of back propagate neural network was 12.6%, and linear regression was 14.7%. Using both acceleration and heart rate for estimation of Tabata energy expenditure is effective, and the prediction effect of back propagate neural network algorithm is better than linear regression, which is more suitable for Tabata energy expenditure monitoring.

Keywords: Tabata training, machine learning, heart rate, acceleration, energy expenditure

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Rana Zeeshan Ahmad,
University of Sialkot, Pakistan
Shubham Mahajan,
Shri Mata Vaishno Devi University, India
Praveen Kumar,
VIT University, India

***Correspondence:**
Qingguo Chen
904350896@qq.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *29 October 2021*
Accepted: *29 November 2021*
Published: *07 February 2022*

Citation:
Yan Y and Chen Q (2022) Energy Expenditure Estimation of Tabata by Combining Acceleration and Heart Rate. Front. Public Health 9:804471. doi: 10.3389/fpubh.2021.804471

INTRODUCTION

According to the recently published World Wide Survey of Fitness Trends, high-intensity interval training (HIIT) has become increasingly popular modes of physical exercise (1). Tabata training is one of the most energetically effective high-intensity intermittent training methods (2).This exercise is relatively inexpensive, usually requires very little equipment, and is effective in promoting physical fitness. Tabata is not only useful to promote fat loss, improve blood pressure, insulin sensitivity, and glucose regulation in a relatively short time, but also enhance sports performances that depend on both the aerobic and anaerobic energy exependiture systems (2, 3). Being able to monitor energy expenditure (EE) portable and accurately during these activities is critical for weight control, physical activity adjustment and fitness goals achievement.

At present, the methods that produce the most accurate measurement of EE is direct calorimetry, indirect calorimetry and doubly labeled water (4), however, one limitation of all three methods is that they are expensive to test, which is not suitable for mass measurement of daily physical activity (5). Both accelerometer and heart rate sensor have long been recognized as the common and cheap wearable devices used for monitoring EE (6, 7). In 2001, Strath et al. (8) proposed a method for predicting EE using accelerated combined heart rate (ACC-HR), and believed that this method is better than the single acceleration (ACC) or heart rate (HR) measurement. So far, other authors have verified that this method predict the EE of some Physical activity (PA) more accurately, such as flat running and uphill running with different intensity (9, 10). In the era of big data, machine learning (ML) has been widely used in people's production and life (11–14). In terms of EE prediction algorithm, back propagate neural network (BPNN) is one of the most common used in machine learning algorithms (15), which has been applied and has significantly improved EE prediction performance in some exercises compared with linear regression (LR).

The main contributions of this paper are:

(1) We analyzed the correlation between ACC and EE, HR and EE in Tabata training.
(2) We established linear regression and back propagate neural network model based on ACC-HR to predict the EE of Tabata training during exercise and intervals.
(3) We evaluate and compare the predictive effectiveness of linear regression and back propagate neural network model.

The rest of the paper is organized as follows. Section 2 gives literature review. Section 3 describes our proposed method in detail. Section 4 presents the test results and analysis. Section 5 discuss the result. Section 6 gives conclusions and Future work.

LITERATURE REVIEW

Existing Wearable Devices

Currently, wearable devices that monitor EE can be divided into five categories: pedometer, accelerometer, heart rate sensor and multi-sensor combination (16). Pedometer is the most convenient motion sensor, and converts the PA of exercisers through the number of walking steps. However, there are significant limitations in EE predicting for slow walking, non-walking, and running (17). The accelerometer can calculate the rate of change of speed within a certain time, and the PA parameters was obtained through body acceleration, then EE is calculated based on the relationship between the known ACC and EE. Unfortunately, the accuracy of energy monitoring is limited by different acceleration wearing positions, additional load on the user and the change of moving surface (18). Heart rate sensor was estimated based on the correlation between HR and oxygen consumption (VO_2) of some physical activities of exercisers, and then the EE was reflected by oxygen intake. But the accuracy of HR monitoring was affected by exercise action, mood and health status (19). A potential and powerful energy expenditure prediction approach is to use ACC and HR simultaneously. ACC values confirm that increased HR is caused by physical activity, which is the basic principle behind the combination of these techniques (16). In addition, the heart rate is linear with the EE of medium and high intensity exercise, which can quantify the daily EE of most people, but it cannot quantify the low intensity physical activity. The deficiency is compensated by acceleration data, which can effectively predict low intensity physical activity (16). In this study, Tabata exercise EE was monitored, including high-intensity exercise and intermittent exercise. The ACC-HR measurements may improve the EE monitoring effect.

Estimation Algorithm

Most of the early research and monitoring EE models are LR models, and the most classical EE linear regression model is Freedsom regression with acceleration parameters (20). Later studies confirmed that a single variable regression could not fit all physical activities, and then some studies developed linear regression models with multiple variables such as acceleration combined with body weight and heart rate combined with body weight for achieving better prediction results (10, 21). However, the ACC, HR, and EE used for some workouts are not linearly or highly correlated. Recently, Morris et al. have shown that ACC is not effective in predicting high-intensity intermittent exercise using the previous LR model (22). Recent sudden developments in the field of machine learning has increased the potential for remote monitoring and diagnostics using data obtained from wearable devices (23). Such as standing balance estimation (24), ECG signal classification (25), and EE prediction (26), etc. Some studies have achieved good accuracy by using BPNN, random forest (RF) and other machine learning algorithms to build models (23, 27). In addition, Montoye et al. (15) has proved that BPNN is better than linear regression for predicting EE in daily physical activity and partial sports. In terms of variable selection, there is inconsistency between machine learning and traditional regression model in the selection of input variables, as shown in **Table 1**. Recently, the researchers verified that BPNN model was constructed with ACC, HR and morphological indicators as input variables, which had high accuracy in predicting walking, running and resistance movement in EE (23). BPNN is the most popularly used Artificial neural network at present, which make up for the difficulty of linear regression in modeling non-linear data or polynomial regression with correlation between data features. For EE prediction of Tabata training, BPNN model may provide a new prediction idea. In this study, acceleration and heart rate data were selected as input variables, and the EE prediction was used by LR and BPNN model.

METHOD

Participants

A sample of 45 participants were included [Female = 23, Age (AG) = 21.04 ± 2.39 years, body height (BH) = 1.67 ± 0.75 m, body weight (BW) = 59.61 ± 8.27 kg, body

Abbreviations: Tabata, Tabata interval training; HR, heart rate; VM, vector magnitude; BPNN, back propagate neural network; LR, line regression; HIFT, high Intensity functional training.

TABLE 1 | Summary of an existing model in EE prediction.

References	Predict activities	Model type	Variable selection
Shaopeng et al. (20)	Daily physical activity, Walk, run	One-variable linear regression	Acceleration
Kuo et al. (10); Caron et al. (21)	Uphill run; walk	Mutiple linear regression	Acceleration, heart rate; morphological indicators acceleration
Montoye et al. (15)	Daily physical activity, walk, run, resistance exercise	BP neural network	Acceleration
O'Driscoll et al. (27)	Stand, walk, run, slope run	Random forest	Acceleration, heart rate, morphological indicators, subject characteristics
Kang et al. (23)	Run, cycling, resistance exercise	BP neural network	Acceleration, heart rate, body temperature

mass index (BMI) = 21.17 ± 2.16 kg/m^2, body fat rate (BFR) = 18.46 ± 5.47%]. Participants were primary recruited from Si Chuan Normal University. Exclusion criteria included the following: (1) Participants had a history of respiratory, cardiovascular, neurobiological, metabolic, or lower extremity related diseases, (2) Participants were given hypnotic drugs, alcohol, caffeine, or nicotine. This study was approved by the Si Chuan Normal University institutional Research Board before participant recruitment. Immediately upon arrival at the exercise research lab, the researchers described the study details to each participant and obtained written informed consent before proceeding with the protocol.

Instrumentation

EE was measured by indirect calorimetry method with gas metabolism analyzer as calibration, and collects data in kilocalorie as the basic unit. During the Tabata test, participants wore a portable gas metabolism analyzer, four wear site accelerometer (dominant hand, non-dominant hand, right hip, right ankle), and a chest strap heart rate band, all set up according to the manufacturer's instructions. The portable gas metabolism analyzer (K4b2, Cosmed) was used as a calibration to measure the EE of Tabata motion, it was calibrated according to the manufacturer's instructions before test. HR was recorded by chest strap heart rate band (Polar wearlink) during exercise, because K4b2 comes with a synchronization sensor from Polar's heart rate band, the heart rate band synchronizes with K42b measurements. ACC was acquired from four accelerometer (Actigraph GT3X+), which is initialized to set the sampling frequency to 30HZ. Prior to the test, participants identified their dominant hand (defined as the hand with which they wrote), and GT3X + was placed on the dominant hand, non-dominant hand, right hip, right ankle for all tests, the instruments was worn as shown in the **Figure 1**. In order to ensure the synchronization of the accelerometer with k4b2 and Polar heart rate band test, calibrate the computer time of the two devices and set the acceleration start time in advance.

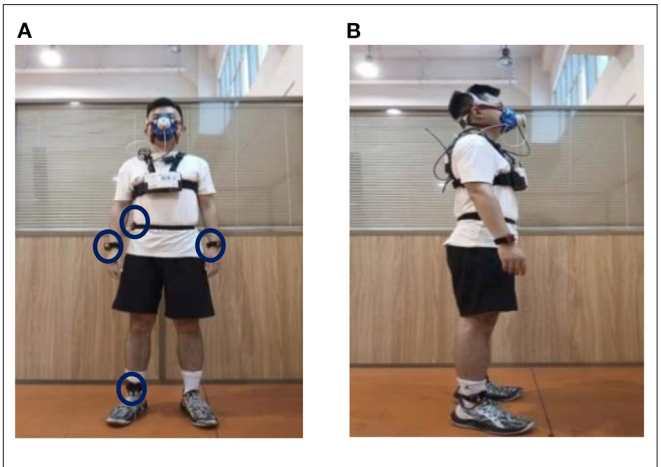


FIGURE 1 | Position of instrument wearing **(A)** Front view and **(B)** Lateral view.

Procedure

Each participant reported to the Sports Research Laboratory for one visit. To avoid discomfort and errors in EE measurement during the test, participants were forbidden to eat 2 h before the test. The testers explained the details of the test to each participant and obtained written informed consent. A PA preparation questionnaire was conducted prior to the test to ensure that participants were healthy and not participating in MVPA contraindications. If the participant answered "yes" to any of the questions on the PA questionnaire, they would have been asked to obtain physician approval before being able to participate in the study; however, whether there are rules in the questionnaire will not be tested. Next, trained research assistants measured height, weight, body fat percentage, and BMI using a human morphological indicator meter (InBody J30). In addition, the age shall be calculated in years according to the date of birth of the resident identity card. The classic Tabata training mode is 20–10 mode (20s-exercise, 10s-interval). In consideration of the simplicity of the exercise movements and the comprehensiveness of the exercise parts, this study adopts the classic Tabata training mode with multiple movement combinations. Participants were asked to learn how to perform the Tabata movements (Jumping jack, Burpees, Running in place high-knees, Side Skaters) and complete each movement 5 times after learning to get familiar with the movements, the exercise process is shown in **Figure 2**, eight movement stages were tested, each consisting of a exercise and an interval. These exercise movements not only can be full-body exercise and follow the principle of simplicity, but also have been proved by studies to achieve high intensity (28), and the intermittent movements are also common. After getting familiar with the action, participants warmed up with a 5-min warm-up in the Keep APP. Once the participant completes warm-up, Polar wearlink, Cosmed K4b2, and Actigraph GT3X fitted to the participant. Sit still for at least 15 min until heart rate stabilizes. At the beginning of the test, participants performed the

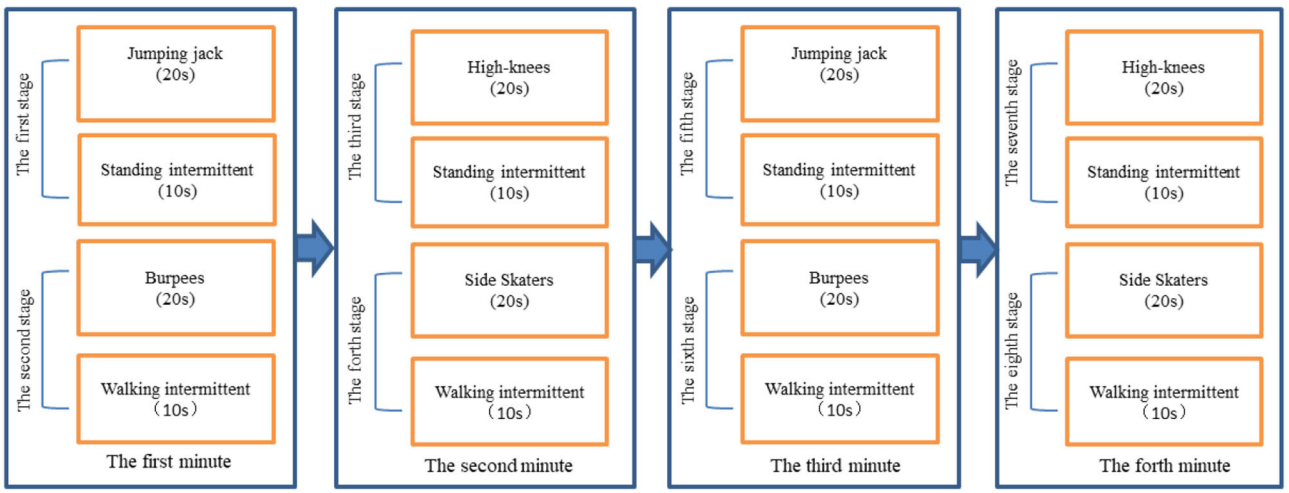


FIGURE 2 | Tabata exercise process.

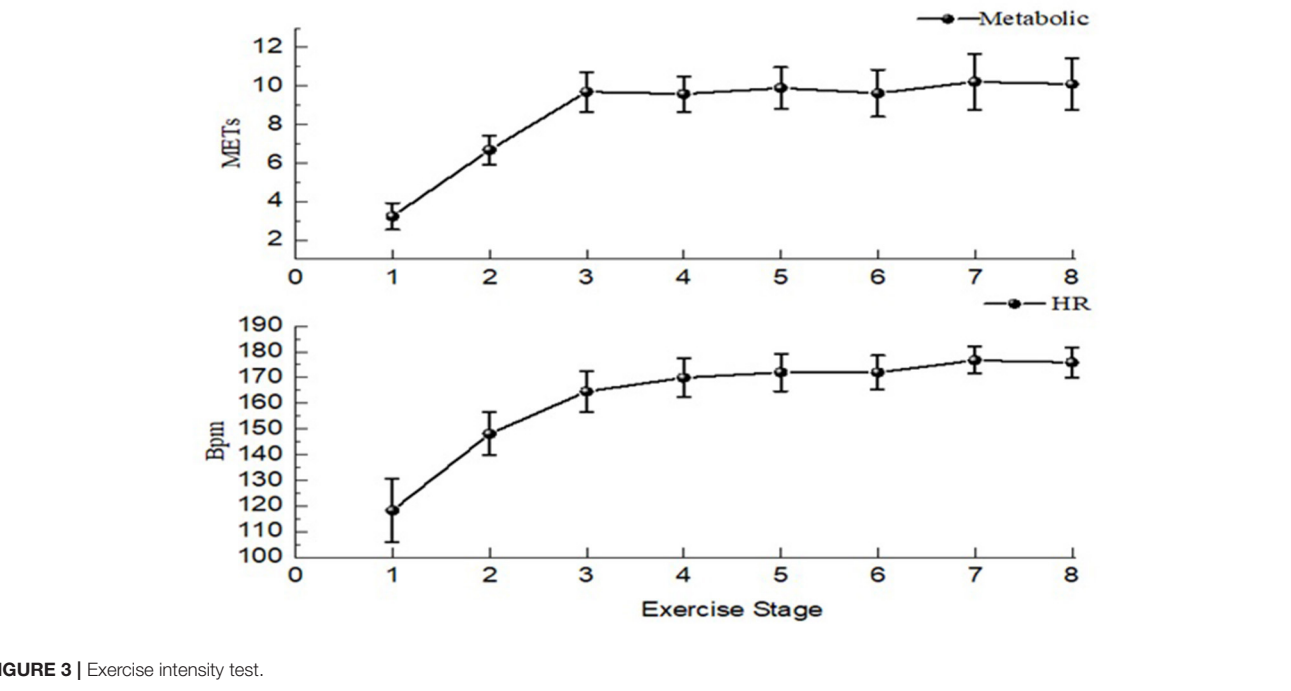


FIGURE 3 | Exercise intensity test.

maximum number of reps prompted by a Tabata video played on a computer. Relax and stretch after completing the test.

Data Processing

Tabata Exercise Intensity Monitoring

To ensure that participants reach the high intensity during the Tabata exercise test, the heart rate of participants and metabolic equivalent (METs) during each stage (exercise + interval) were collected during the test (**Figure 3**). In this study, the average HR of exercise was 165.46 ± 5.39 for men and 163.88 ± 6.88 for women, reaching 80% of the maximum heart rate. As the same time, the average of METs was 8.84 ± 0.89 for men and 8.39 ± 0.93 for women. Refer to ACSM for exercise categories and physical activity intensity (29), the exercise intensity of this test reached high intensity, Maximum heart rate 183.43 ± 7.09, almost 90% of maximum heart rate, which was in line with the exercise intensity required by Tabata exercise.

Variable Selection

The triaxial acceleration values are combined into vector acceleration, the vector magnitude (VM) can been seen in Equation 1 (the X, Y, Z is the acceleration of the three axes, N

means in unit time interval). Data of acceleration, heart rate and EE are intercepted in 10 and 30s.

$$\mathbf{VM}=\frac{\sum_{\mathbf{i=0}}^{\mathbf{N-1}}\sqrt{x_i^2+y_i^2+y_i^2}}{\mathbf{N}} \quad (1)$$

Given the difference in EE predictions between exercise and interval, the selection of linear regression and neural network variables was based on previous studies (20, 23). EE is predicted in kcal/10s. Pre-selected variables during exercise included: human morphological index, the unit of 10s corresponds to heart rate during the exercise and the unit of count/10s corresponds to the VM value of each wearing part during the exercise. Since the interval is 10 s, the VM value is the minimum or "0" during the standing interval, and the heart rate during exercise is related to the EE during the interval (31). The pre-selected variable for predicting the intermittent EE includes: human morphological, VM value during 0–10s exercise, VM value during 10–20s exercise, VM value during 0–20s exercise, VM value during entire stage, HR value during 0–10s exercise, HR value during 10–20s exercise, HR value during 0–20s exercise, HR value during entire stage, HR value in Intermittent. Pearson correlation analysis was conducted between pre-selected variables and EE in exercise and interval period, respectively. The absolute value of correlation coefficient ($|r|$) compared the correlation between each variable and EE.

Model Development

In order to build LR model, the pre-selected variables of the exercise period and the intermittent period were put into stepwise linear regression, respectively. Then the equations with significance *P*-value of variables <0.05 were selected. Finally, LR model of exercise period and intermittent period were constructed.

BP neural network is a multi-layer error feedback neural network, which consists of input layer, hidden layer and output layer. Each node represents a neuron, the upper node and the lower node are connected by weight, and the nodes between layers are fully interconnected. Before input variable selection, the temporal characteristics of the four accelerometers were extracted in the non-overlapping windows of 10 and 30 s, and the heart rate (HR) in units of 10 and 30 s was obtained. Input layer variables are selected based on the degree of correlation between EE and the variable, which can include gender, age, weight, height, average heart rate, mean, standard deviation, cova-riance of adjacent windows of data, minimum, maximum, the 10, 25, 50, 75, and 90th percentiles of the vectorial magnitude acceleration. And these variables have been proved to be effective machine learning EE prediction model can be built method (27). In order to improve the speed of convergence, in this study, momentum—learning rate adaptive adjustment algorithm was used to optimize BPNN. The number of hidden nodes was determined by trial and error method. The formula of hidden layer (32) used the Equation 2, "n_i" is the number of hidden nodes, "no" is the number of input nodes, and "c" is the number of output nodes. After the number of hidden node is determined, the number of nodes is increased step by step until the number of nodes with the lowest Root mean square error (RMSE) is found. The activation functions of the hidden layer and output layer are determined as logsig logs-type transfer functions. Sigmoid type differentiable function is strictly incremental, with solid theoretical basis, rigorous derivation process, beautiful symmetry of the obtained formula, and strong non-linear fitting ability (33), the formula is the Equation 3. The BPNN model was constructed by the variables of neural network model through Matlab software (R2018b).

$$n=\sqrt{n_i+n_0}+c \quad (2)$$

$$Log\ \mathbf{sig}\ (\mathbf{x})=\frac{1}{1+\mathbf{e}^{-\mathbf{x}}} \quad (3)$$

$$\mathbf{RMSE}=\sqrt{\frac{1}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\hat{\mathbf{y}}_\mathbf{i}-\mathbf{y}_\mathbf{i})^2} \quad (4)$$

Statistical Analysis

Random grouping was used in which all available data are split into training and testing, at a ratio of 3:1. After grouping, the predicted values were obtained by substituting the test group data into the prediction model constructed by the training. Using Bland-Altman plot to verify the difference between the predicted and measured values, the model is effective when 95% points fall between ± 1.96 SD of the difference between predicted value and measured value. Mean percentage absolute error (MAPE) and RMSE was used to compare the accuracy of prediction between models, the specific formula is the Equations 4 and 5, "$\hat{y}_i$" is the actual measured value, "y_i" is the predicted value.

$$MAPE=\left(\frac{\mathbf{100}}{n}\right)\sum_{\mathbf{i=1}}^{\mathbf{n}}\left|\frac{\hat{\mathbf{y}}\mathbf{i}-\mathbf{y}\mathbf{i}}{\mathbf{y}\mathbf{i}}\right| \quad (5)$$

RESULT

Correlation Analysis

According to Pearson correlation analysis, we compared the correlation between each variables and EE (**Figure 4**). The heart rate band data is the significant correlated with exercise and interval EE. In the various indicators of heart rate, the highest correlation with EE during exercise period is HR value during exercise ($r = 0.758$), the highest correlation with EE during intermittent period is HR value during 0–10s exercise ($r = 0.655$). All the morphologic indicators are significantly correlated with EE, and the highest correlation between anthropometric index and EE in all stages is body weight ($r = 0.331$, $r = 0.581$). Although, the correlation between EE and ACC of multiple wear sites in all stages is significant, the highest correlation variables with EE during exercise and interval period is only moderately correlated, they are right ankle VM value during exercise and right ankle VM value during entire stage ($r = 0.352$, $r = 0.308$). Only the right ankle and hand acceleration are more than 0.300 correlated with EE during the interval. The correlation between the four wearing parts ACC and EE is different in both the exercise and the interval period, and the correlation of ankle is the largest. The measurement indexes of wearable devices with the

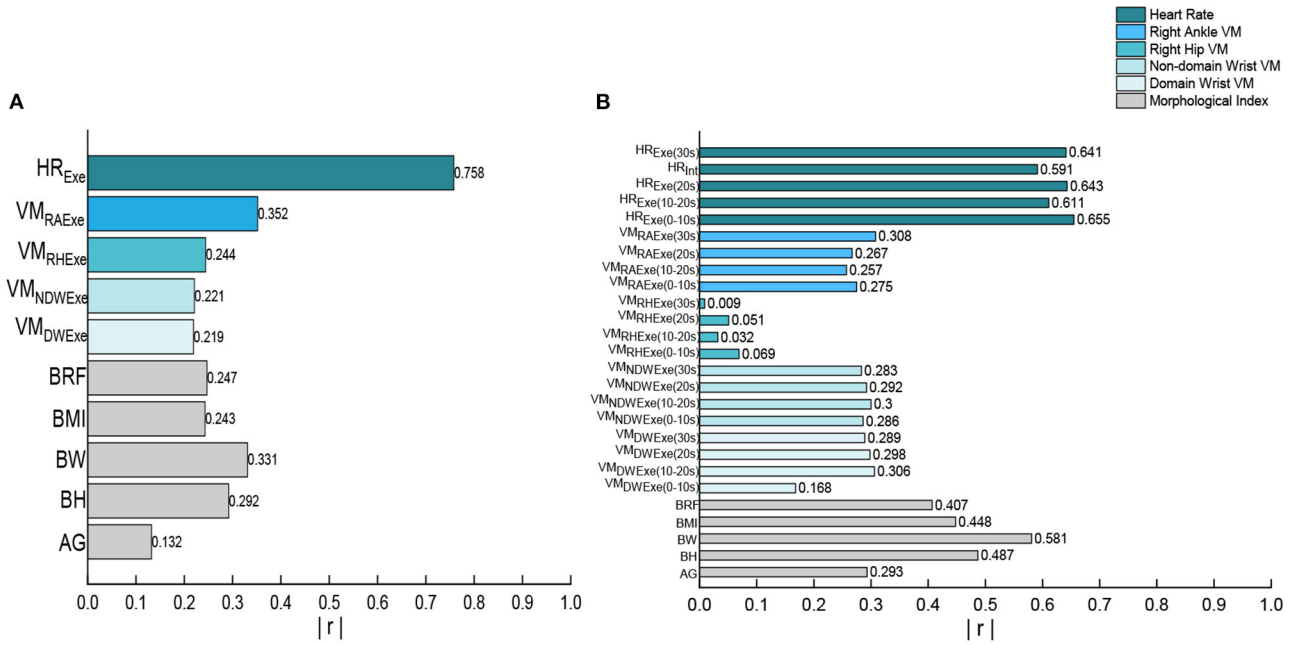


FIGURE 4 | (A) Correlation analysis of exercise period; **(B)** Correlation analysis of interval period.

greatest correlation were analyzed, as shown in **Figure 5**. There are linear correlation between heart rate measurement indexes and EE in the whole Tabata training, while all the accelerometer indexes are weakly linear correlated with EE.

Linear Regression

The variables were put into linear stepwise regression, and the LR model was established as shown in **Table 2**. right ankle VM value during exercise, HR value during exercise and body weight were took as independent variables to build LR model of EE during exercise period. VM value during entire stage, HR value during 0–10s exercise and body weight were took as independent variables to build LR model of EE during interval period. The significance *P*-values of these independent variables in the LR model are all <0.05, and they are all valid variables. The r^2 value of the EE prediction model in exercise period is 0.710, and the r^2 value of the EE prediction model in interval period is 0.730. The significance of the model is <0.05, indicating that the model is valid.

BP Neural Network

Based on the correlation between the acceleration data of each position and EE during all stages, this study decided to take the right ankle VM value during exercise as the input variable of the BPNN model during the exercise period, and the right ankle VM value during entire stage was input as the input variable of the BPNN model in intermittent period. HR in exercise period was selected as the input variable value of the neural network model during the exercise period, and HR value during 0–10s exercise was taken as the input variable of the BPNN model during the exercise period. Height, weight and gender were input into the two models as the anthropometric index and basic information of participants. After using trial and error method, we determined 4–15 hidden layers as an attempt, and finally got the hidden layers of the neural network as shown in **Figure 6**. The number of hidden layers of model in exercise period was 6, and that of model in intermittent period was 10.

Finally, the neural network model was determined as 14-6-1 in the exercise period and 14-10-1 in the interval period. The upper limit of network iteration of neural network model is 5,000 times, initial learning rate is 0.05, momentum constant is 0.9 and error rate is 0.001, as shown in **Figure 7**. "w1" and "w2" in turn represent input and sum weight vectors, "Si" means the sum of the input layers in the implicit cell, "U" indicates that the input activation function is hidden. The training group was substituted into the model for training and BPNN was finally obtained. After calculation, the time complexity of BPNN model during exercise and interval period is 20139840 and 27643680, respectively. The r^2 value of the exercise model is 0.813, the r^2 value of the intermittent model is 0.816, which indicates the model is valid.

Predictive Accuracy of Models

The variable data of the test group was substituted into each prediction model and the predicted values of each model were obtained. Then the predicted EE values and measured EE values of 15 subjects in the validation group were intercepted in 10-s increments, each subject will have 24 predicted values. The whole verification group was substituted into each model to obtain 24*15 predicted values. The forecast and measured graph was presented in **Figure 8**.

Bland-Altman Plot statistical method with MedCalc software was used to analyze the consistency of model prediction effect, as shown in **Figure 7**. The difference between the measured EE

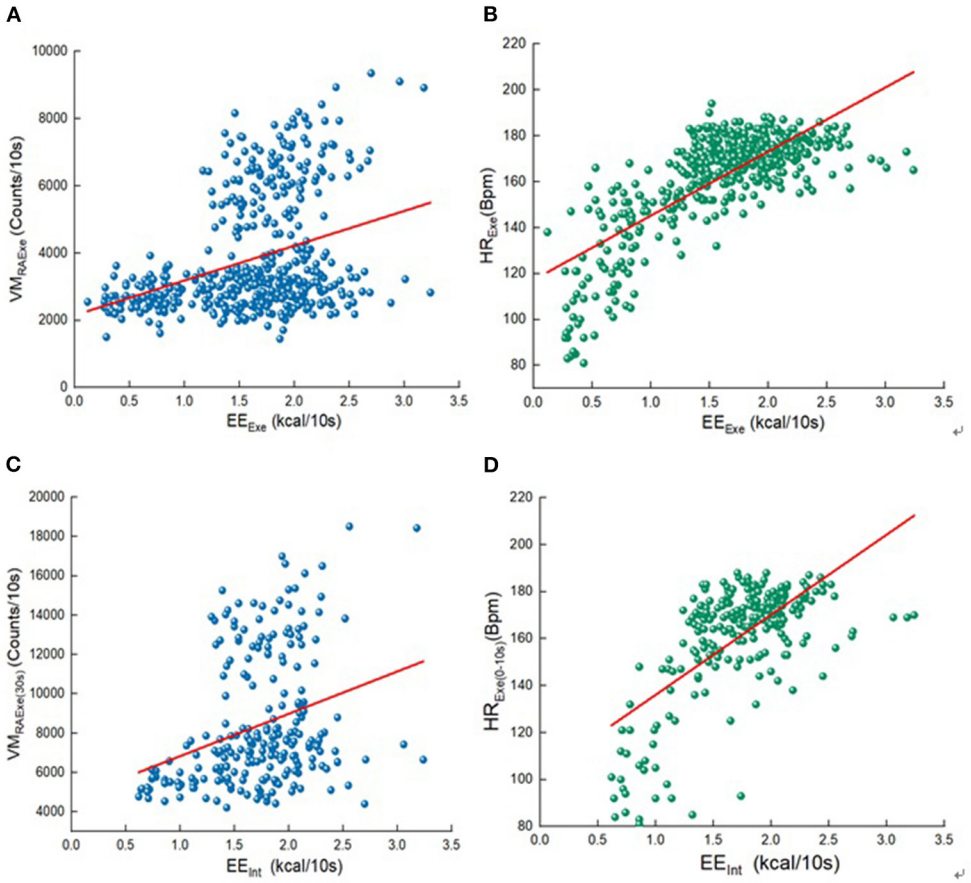


FIGURE 5 | Variables and EE scatter diagram. **(A)** the unit of 10s corresponds to VM during the exercise and EE_{Exe}; **(B)** the unit of 10s corresponds to HR during the exercise and EE_{Exe}; **(C)** VM value during 30s exercise and EE_{Int}; **(D)** HR value during 0–10s exercise and EE_{Int}.

TABLE 2 | EE prediction model of Tabata linear regression.

Use conditions	Model	r^2
Exercise	$EE_{Exe} = 0.000044*VM_{RAExe} + 0.193*HR_{Exe} + 0.23*BW-3.05$	0.71
Interval	$EE_{Int} = 0.000011*VM_{RA(30s)} + 0.0116*HR_{Exe(0-10s)} + 0.030*BW-1.99$	0.73

value of K4b2 and the predicted EE value of each model was used as the Y-axis, and the mean value of the predicted EE value of each model and the measured K4b2 was used as the X-axis. **Figure 9** showed that 95% of the scatter points of the two prediction models were within the range of ± 1.96 SD during the exercise period and interval period, they all have good consistency.

The predicted and measured values of the validation group were substituted into the MAPE formula for calculation, and the MAPE of the two models in the exercise period and the intermittent period were compared. The MAPE of LR model is 16.95% in EE prediction during exercise period, which is 2.65% higher than BPNN model. In the intermittent EE prediction, the MAPE of BPNN model is 9.28%, which is lower than LR model. At the same time, the error dispersion degree of LR model is larger, and its standard error of MAPE is 1.70% and 1.22%, which proves that the stability of LR prediction effect is lower than that of BPNN (**Figure 10**).

In order to compare the prediction effect of the Tabata training during each movement stage, the two models were compared in each exercise stage, as shown in **Figure 11**. In the LR model, the MAPE of three exercise stages is >15%, while the BPNN model has two. In addition, the MAPE and RMSE predicted by tow EE model in entire Tabata training was calculated. The MAPE of BPNN model was 12.6%, which is 2% lower than that of LR model, and the RMSE of LR model was 0.238, which is 0.011 higher than that of BPNN model. Both LR and BPNN models can effectively predict the EE of Tabata, but BPNN model has better prediction performance than the LR model regardless of the comparison of each part or the whole movement.

DISCUSSION

In this study, the correlation of pre-selected variables was analyzed. The highest correlation with the EE in the exercise and the intermittent period is HR, which basically shows a linear relationship with the EE. The linear correlation between HR and EE in high-intensity exercise can also be confirmed in Tabata training (16). Bazuelo et al. (30) has shown that intermittent heart rate is related to intermittent EE. But this study found

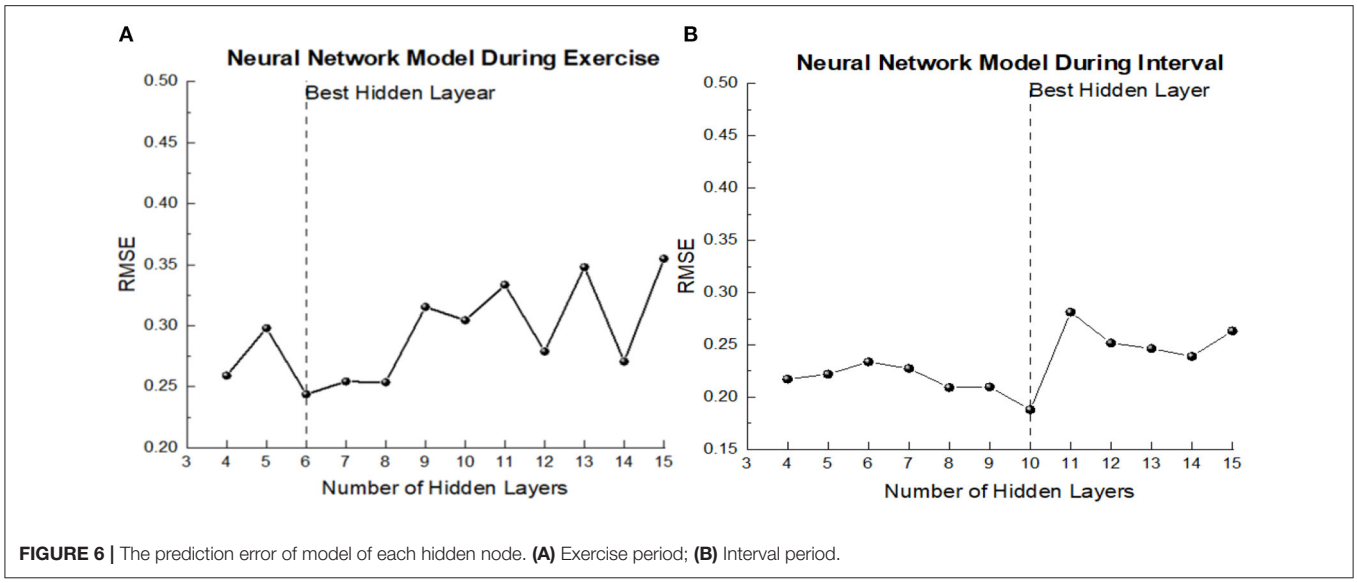


FIGURE 6 | The prediction error of model of each hidden node. **(A)** Exercise period; **(B)** Interval period.

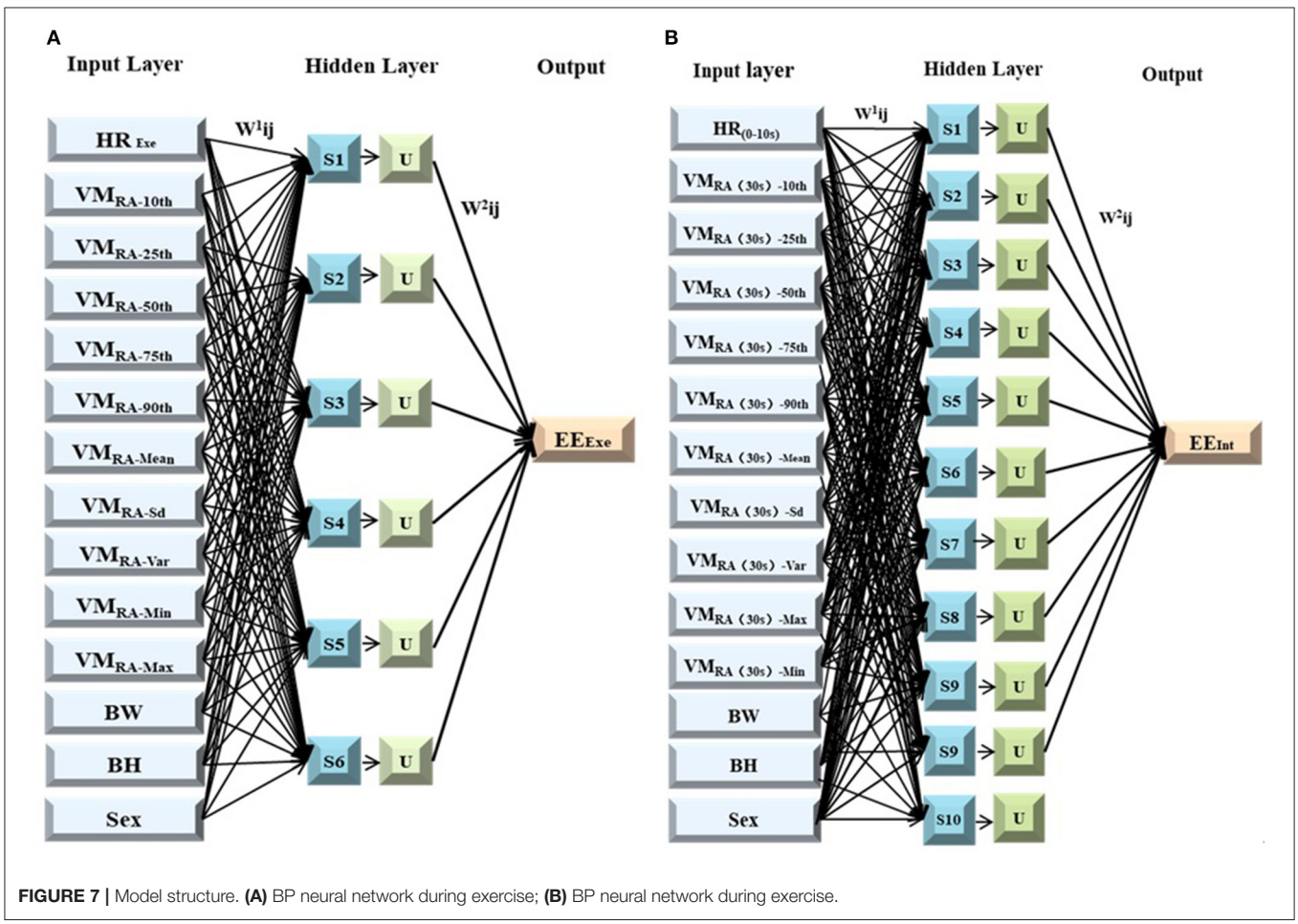


FIGURE 7 | Model structure. **(A)** BP neural network during exercise; **(B)** BP neural network during exercise.

that intermittent EE not only correlate with intermittent heart rate, but also with exercise heart rate, the highest correlation with intermittent EE was the HR value during 0–10s exercise, and the correlation was basically linear. Meanwhile Dugas et al. show the similar result (31), the exercise they tested was set-up exercise (45s-exercise, 15s-interval), the significant correlation of

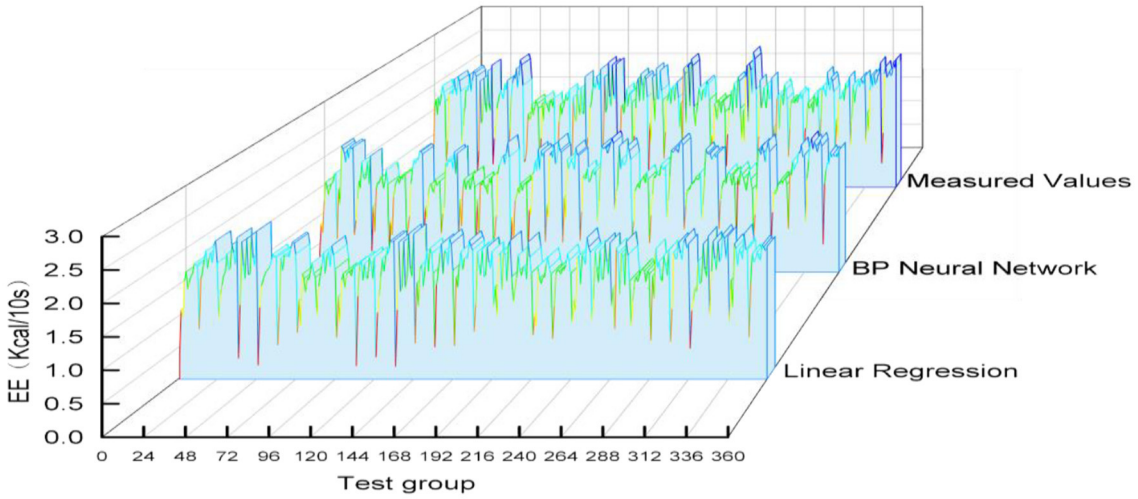


FIGURE 8 | Predicted values and Measured values.

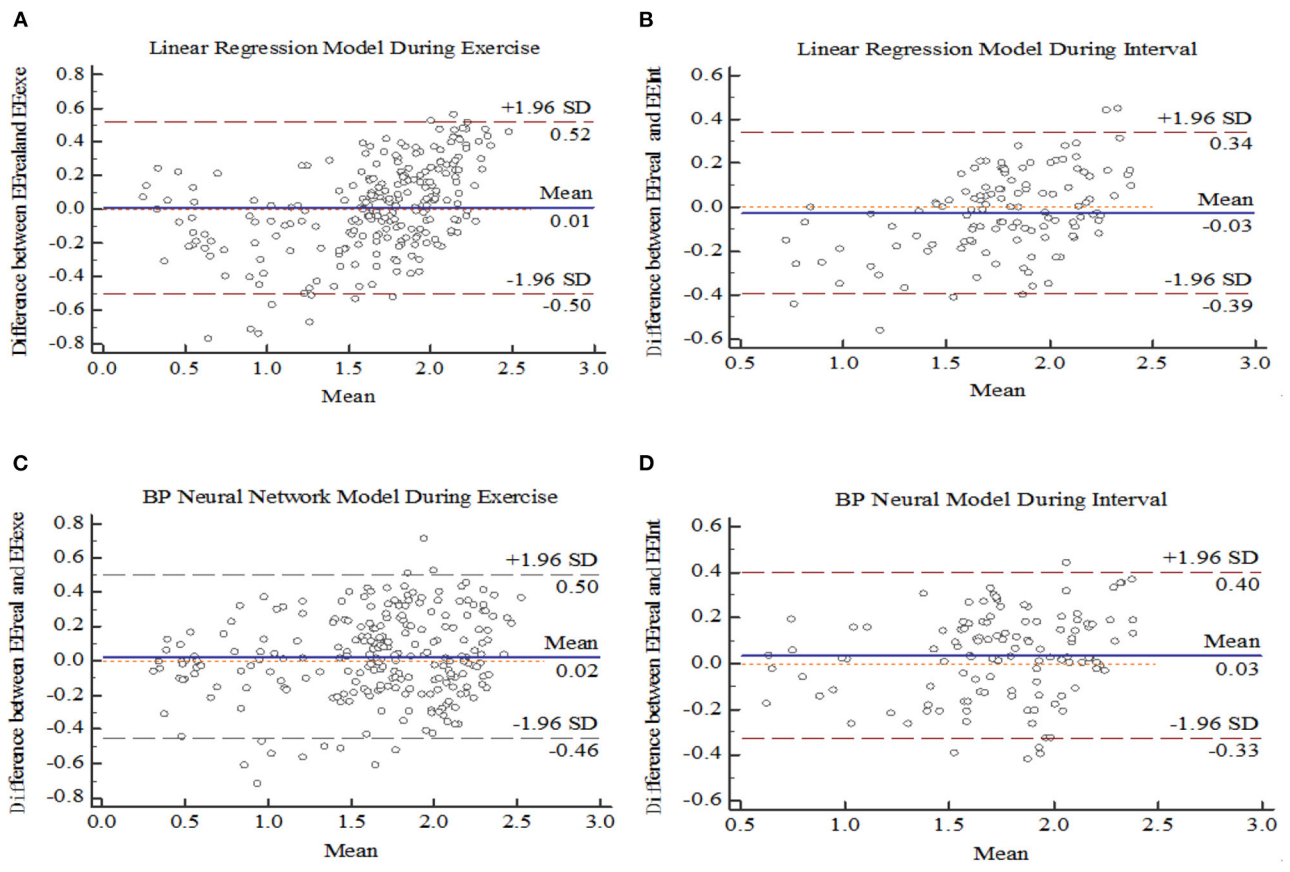


FIGURE 9 | Bland Altman plot for models. **(A)** Linear regression during Exercise. **(B)** Linear regression during interval. **(C)** BP neural network during Exercise **(D)** BP neural network during interval.

intermittent EE is exercise HR 1 min before the interval. Due to a temporal dissociation be-tween HR and VO2 (EE) during intermittent-type activity, the relationship between resting heart rate and oxygen uptake was not strong during partial exercise intervals (34), Therefore, the HR value during 0–10s exercise was selected as the input variable in this study. At present, there are

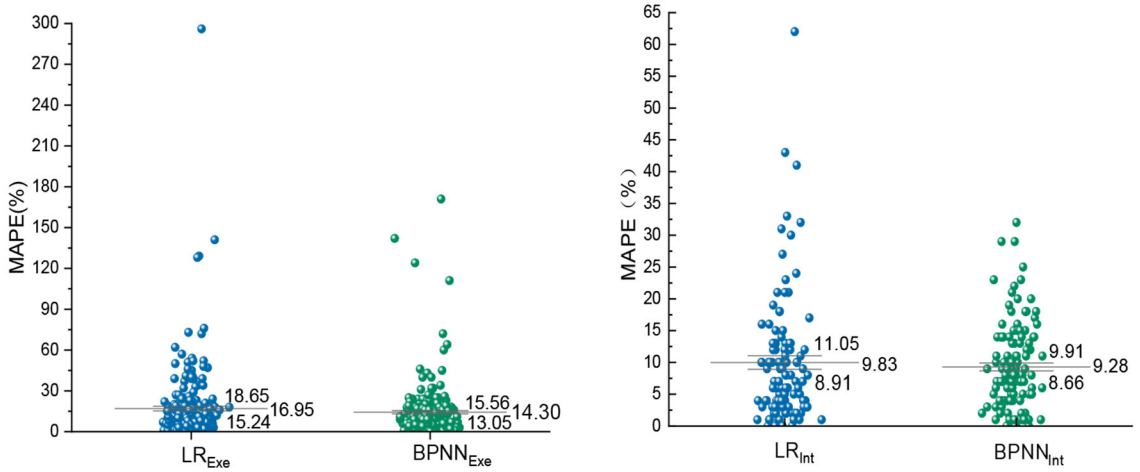


FIGURE 10 | Prediction error of EE during exercise and interval period.

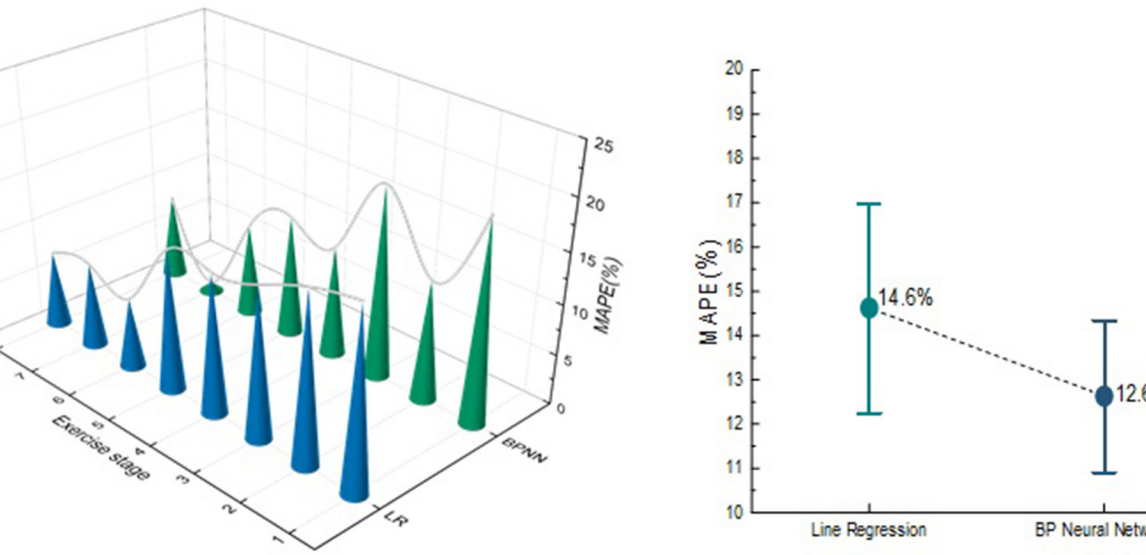


FIGURE 11 | MAPE comparison of the EE of total exercise.

few studies on the correlation between exercise heart rate and intermittent energy, and the correlation between other exercises and intermittent energy expenditure is not clear, which needs to be verified in the future.

Throughout the Tabata training, the ACC has a moderate correlation with EE, showing a weak linear correlation with EE, and the ankle ACC is the highest correlation among all the wearing parts. Tabata tested in this study included high-intensity and complex movements. In this exercise mode, the ACC values of each part of the body are vary a lot and do not show periodic changes, and the different parts of the body show different physical activity. This is similar to the study by Morris et al. (22), who validated smart devices with built-in accelerometers worn on the hip and wrist. The high intensity intermittent motion was carried out, the prediction efficiency of the acceleration values of the two wear parts is not good after substituting into the existing linear model. Although some studies suggest that accelerometers worn on the hip can better predict EE, most of these studies tested continuous periodic motion (35). Some studies have also suggested that ankle is the most suitable wearing part for sports with high speed and intensity, because the acceleration signal vector of human motion changes with different postures (36). Therefore, the ankle acceleration can be used as a prediction variable in EE prediction in this study.

Some studies have also shown the correlation between EE and anthropometric indexes (37). In this study, the correlation between body weight and EE was the highest among anthropometric indexes. Therefore, body weight is also considered a variable of EE prediction.

After the correlation analysis of the pre-selected variables, the LR model was constructed with the heart rate, the right ankle VM value and body weight as independent variables through stepwise regression of the predictive variables. Referring to input variables selected in past studies, BPNN is constructed by combining relevant variables significantly related to EE with subject characteristics information.

In terms of model establishment, Bland-Altman method is used to test that LR and BPNN model are consistent. The feasibility of predicting EE with heart rate combined accelerometer was confirmed, which was consistent with the results of previous studies by Brage and Crouter (38, 39). A recent study evaluated the effectiveness of accelerometers in predicting the EE of High Intensity Functional interval Training (HIFT), which is similar to the Tabata model. But, the effect of accelerometer alone on the EE prediction by linear regression model is not ideal, and the MAPE of energy expenditure in the whole movement process is more than 15%, and the errors in predicting EE during exercise and interval period have not been specifically evaluated (22). In accordance to the statement made by Lee et al. (40) that the reasonable error of energy expenditure prediction is 10–15%. Compared with the recent study, the overall prediction errors of the two ACC-HR models constructed in this study are both lower than 15%, which has improved the prediction effect in EE prediction of high intensity intermittent exercise. By comparing LR and BPNN model, the MAPE of BPNN model is higher than 15% in two stages, while the MAPE in LR model is higher than 15% in three stages. In addition, the RMSE and MAPE predicted by BPNN in the whole exercise process were lower than those predicted by LR model (0.227 and 12.6%), which proved that BPNN model based on machine learning algorithm in the era of big data could provide exercisers with more accurate prediction of EE than LR, which was similar with the result of Montoya et al. (15).

LIMITATION

Firstly, Limitations of test contents, this study selected classic and representative Tabata exercise pattern and movement for testing. Although the model cannot predict the EE of all Tabata exercises, it can be used to predict EE in the classical pattern. Future research can further enrich the testing of Tabata exercises and modes, so as to improve the accuracy of Tabata EE prediction. In addition, Limitations of the applicability, the subjects selected were 45 young adults. Therefore, the generalization of the models established in this study to other age group further research and confirmation.

CONCLUSION AND FUTURE WORK

In this paper, a linear regression and neural network model were constructed to predict the EE of Tabata training using accelerometer combined with heart rate data. The two models have good consistency by bland-Alterman test, and the prediction performance is evaluated by MAPE and RMSE, which proves that the error of back propagate neural network is lower than linear regression. Overall, the empirical results showed that: (i) the effectiveness of acceleration combined with heart rate data in predicting Tabata exercise EE. (ii) Back propagate neural network model constructed by acceleration combined with heart rate has higher prediction accuracy than linear regression model, reflecting the feasibility of machine learning algorithm in predicting EE of complex intermittent exercise. The ACC-HR Back propagate neural network model provides exercisers with a portable and more accurate tool for predicting Tabata EE, so as to help exercisers adjust their physical activities and set up their exercise plans, thus improving their exercise benefits. In the future, more novel and useful machine learning algorithms will be used to predict the EE of high intensity interval exercise. In addition, this study will test more Tabata exercise patterns and movements to more fully and accurately predict Tabata exercise EE.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Chengcheng Cai, Sport Laboratory, College of Physical Education, Sichuan Normal University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YY participated in research concepts, design, subject data collection, manuscript writing, and chart making. QC contributed to the research concept and made critical revisions to the research design and manuscript. All authors contributed to the paper, the manuscript was constantly revised, and the authors unanimously agreed on the version submitted.

FUNDING

The work was supported by the Physical Education Institute of Sichuan Normal University, which provided the venue for the testing.

ACKNOWLEDGMENTS

The authors would like to thank Junsong Pu, Yubing Zhou and Junrong Zhang for their assistance with test data collection.

REFERENCES

1. Thompson WR. Worldwide survey of fitness trends for 2019. *ACSM Health Fitness J.* (2018) 22:10–7. doi: 10.1249/fit.0000000000000438
2. Tabata I. Tabata training: one of the most energetically effective high-intensity intermittent training methods. *J Physiol Sci.* (2019) 69:559–72. doi: 10.1007/s12576-019-00676-7
3. Olson M. TABATA: It'sa HIIT. *ACSM Health Fit J.* (2014) 18:17–24. doi: 10.1249/fit.0000000000000065
4. Altini M, Penders J, Vullers R, Amft O. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. *IEEE J Biomed Health.* (2014) 19:219–26. doi: 10.1109/jbhi.2014.2313039
5. Ndahimana D, Kim E-K. Measurement methods for physical activity and energy expenditure: a review. *Clin Nutr.* (2017) 6:68–80. doi: 10.7762/cnr.2017.6.2.68
6. Tikkanen O, KäRKKäINEN S, Haakana P, Kallinen M, Pullinen T, Finni T. EMG, heart rate, and accelerometer as estimators of energy expenditure in locomotion. *Med Sci Sports Exerc.* (2014) 46:1831–9. doi: 10.1249/mss.0000000000000298
7. Vernillo G, Savoldelli A, Pellegrini B, Schena F. Validity of the SenseWear Armband to assess energy expenditure in graded walking. *Ment Health Phys Act.* (2015) 12:178–83. doi: 10.1123/jpah.2013-0437
8. Strath SJ, Bassett DR, Swartz AM, Thompson DL. Simultaneous heart rate-motion sensor technique to estimate energy expenditure. *Med Sci Sports Exerc.* (2001) 33:2118–23. doi: 10.1097/00005768-200112000-00022
9. Ho C-S, Chang C-H, Hsu Y-J, Tu Y-T, Li F, Jhang W-L, et al. feasibility of the energy expenditure prediction for athletes and non-athletes from ankle-mounted accelerometer and heart rate monitor. *Sci Rep.* (2020) 10:1–9. doi: 10.1038/s41598-020-65713-7
10. Kuo TB, Li J-Y, Chen C-Y, Lin Y-C, Tsai M-W, Lin S-P, et al. Influence of accelerometer placement and/or heart rate on energy expenditure prediction during uphill exercise. *J Mot Behav.* (2018) 50:127–33. doi: 10.1080/00222895.2017.1306481
11. Gadekallu TR, Rajput DS, Reddy M, Lakshmanna K, Bhattacharya S, Singh S, et al. A novel PCA–whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *J Real Time Image PR.* (2021) 18:1383–96. doi: 10.1007/s11554-020-00987-8
12. Dhanamjayulu C, Nizhal U, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Proc.* (2021). doi: 10.1049/ipr2.12222
13. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors. *J Real Time Image Proc.* (2021) 18:1773–85. doi: 10.1007/s11554-021-01122-x
14. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst.* (2021) 28:1–15. doi: 10.1007/s00530-021-00774-w
15. Montoye AH, Begum M, Henning Z, Pfeiffer KA. Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiol Meas.* (2017) 38:343. doi: 10.1088/1361-6579/38/2/343
16. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. *Front Nutr.* (2014) 1:5. doi: 10.3389/fnut.2014.00005
17. Abel MG, Peritore N, Shapiro R, Mullineaux DR, Rodriguez K, Hannon JC. A comprehensive evaluation of motion sensor step-counting error. *Appl Physiol Nutr ME.* (2011) 36:166–70. doi: 10.1139/h10-095
18. Corder K, Brage S, Wareham NJ, Ekelund U. Comparison of PAEE from combined and separate heart rate and movement models in children. *Med Sci Sports Exerc.* (2005) 37:1761–7. doi: 10.1249/01.mss.0000176466.78408.cc
19. Livingstone M, Robson PJ, Totton M. Energy expenditure by heart rate in children: an evaluation of calibration techniques. *Med Sci Sports Exerc.* (2000) 32:1513–9. doi: 10.1097/00005768-200008000-00022
20. Liu S, Gao R, Freedson P. Computational methods for estimating energy expenditure in human physical activities. *Med Sci Sports Exerc.* (2012) 44:2138. doi: 10.1249/mss.0b013e31825e825a
21. Caron N, Peyrot N, Caderby T, Verkindt C, Dalleau G. Estimating energy expenditure from accelerometer data in healthy adults and patients with type 2 diabetes. *Exp Gerontol.* (2020) 134:110894. doi: 10.1016/j.exger.2020.110894
22. Morris CE, Wessel PA, Tinius RA, Schafer MA, Maples JM. Validity of activity trackers in estimating energy expenditure during high-intensity functional training. *Res Q Exerc Sport.* (2019) 90:377–84. doi: 10.1080/02701367.2019.1603989
23. Kang KH, Shin SH, Jung J, Kim YT. Estimation of a physical activity energy expenditure with a patch-type sensor module using artificial neural network. *Concurr Comp Pract E.* (2021) 33:e5455. doi: 10.1002/cpe.5455
24. Lattanzi E, Freschi V. Evaluation of human standing balance using wearable inertial sensors: a machine learning approach. *Eng Appl Artif Intell.* (2020) 94:103812. doi: 10.1016/j.engappai.2020.103812
25. Hua J, Xu Y, Tang J, Liu J, Zhang J. ECG heartbeat classification in compressive domain for wearable devices. *J Syst Architect.* (2020) 104:101687. doi: 10.1016/j.sysarc.2019.101687
26. Xiao L, Wu K, Tian X, Luo J. Activity-specific caloric expenditure estimation from kinetic energy harvesting in wearable devices. *Perv Mobile Comput.* (2020) 67:101185. doi: 10.1016/j.pmcj.2020.101185
27. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. Improving energy expenditure estimates from wearable devices: a machine learning approach. *J Sport Sci.* (2020) 38:1496–505. doi: 10.1080/02640414.2020.1746088
28. Emberts T, Porcari J, Dobers-Tein S, Steffen J, Foster C. Exercise intensity and energy expenditure of a tabata workout. *J Sci Med Sport.* (2013) 12:612. doi: 10.1249/01.mss.0000762452.62307.d8
29. Medicine ACoS. *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription*. Philadelphia, PA: Lippincott Williams & Wilkins (2012) .
30. Bazuelo-Ruiz B, de Rosario H, Durá-Gil JV. Validation of oxygen consumption prediction equations and new formulas for interval training. *Apunt Educ Fisy Deport.* (2021) 143:9–16. doi: 10.5672/apunts.2014-0983.es.(2021/1).143.02
31. Dugas LR, Van Der Merwe L, Odendaal H, Noakes TD, Lambert EV. A novel energy expenditure prediction equation for intermittent physical activity. *Med Sci Sport Exer.* (2005) 37:2154–61. doi: 10.1249/01.mss.0000177742.12931.50
32. West D. Neural network credit scoring models. *Comput Oper Res.* (2000) 27:1131–52. doi: 10.1016/s0305-0548(99)00149-5
33. Yılmaz S, Demircioglu C, Akin S. Application of artificial neural networks to optimum bit selection. *Comput Geosci.* (2002) 28:261–9. doi: 10.1016/s0098-3004(01)00071-1
34. Bot S, Hollander A. The relationship between heart rate and oxygen uptake during non-steady state exercise. *Ergo.* (2000) 43:1578–92. doi: 10.1080/001401300750004005
35. Leenders N, Nelson T, Sherman W. Ability of different physical activity monitors to detect movement during treadmill walking. *Int J Nanomedicine.* (2003) 24:43–50. doi: 10.1055/s-2003-37196
36. Jung Y-S, Park R-W, Joo N-S. Different location of triaxial accelerometer and different energy expenditures. *Yonsei Med J.* (2014) 55:1145–51. doi: 10.3349/ymj.2014.55.4.1145
37. LeCheminant JD, Heden T, Smith J, Covington NK. Comparison of energy expenditure, economy, and pedometer counts between normal weight and overweight or obese women during a walking and jogging activity. *Eur J Appl Physiol.* (2009) 106:675–82. doi: 10.1007/s00421-009-1059-9
38. Brage S, Brage N, Franks PW, Ekelund U, Wong M-Y, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured

physical activity energy expenditure. *Eur J Appl Physiol.* (2004) 96:343–51. doi: 10.1152/japplphysiol.00703.2003
39. Crouter SE, Churilla JR, Bassett DR. Accuracy of the actiheart for the assessment of energy expenditure in adults. *Eur J Clin Nutr.* (2008) 62:704–11. doi: 10.1038/sj.ejcn.1602766
40. Lee J-M, Kim Y-W, Welk GJ. TRACK IT: validity and utility of consumer-based physical activity monitors. *ACSMs Health Fit J.* (2014) 18:16–21. doi: 10.1249/fit.0000000000000051

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yan and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 15 February 2022
doi: 10.3389/fpubh.2021.797877

A Novel Extra Tree Ensemble Optimized DL Framework (ETEODL) for Early Detection of Diabetes

Monika Arya[1], Hanumat Sastry G[2], Anand Motwani[3], Sunil Kumar[2] and Atef Zaguia[4]*

[1] Department of Computer Science and Engineering, Bhilai Institute of Technology, Durg, India, [2] School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India, [3] School of Computing Science and Engineering, VIT Bhopal University, Sehore, India, [4] Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

Diabetes has been recognized as a global medical problem for more than half a century. Patients with diabetes can benefit from the Internet of Things (IoT) devices such as continuous glucose monitoring (CGM), intelligent pens, and similar devices. Smart devices generate continuous data streams that must be processed in real-time to benefit the users. The amount of medical data collected is vast and heterogeneous since it is gathered from various sources. An accurate diagnosis can be achieved through a variety of scientific and medical techniques. It is necessary to process this streaming data faster to obtain relevant and significant knowledge. Recently, the research has concentrated on improving the prediction model's performance by using ensemble-based and Deep Learning (DL) approaches. However, the performance of the DL model can degrade due to overfitting. This paper proposes the Extra-Tree Ensemble feature selection technique to reduce the input feature space with DL (ETEODL), a predictive framework to predict the likelihood of diabetes. In the proposed work, dropout layers follow the hidden layers of the DL model to prevent overfitting. This research utilized a dataset from the UCI Machine learning (ML) repository for an Early-stage prediction of diabetes. The proposed scheme results have been compared with state-of-the-art ML algorithms, and the comparison validates the effectiveness of the predictive framework. This proposed work, which outperforms the other selected classifiers, achieves a 97.38 per cent accuracy rate. F1-Score, precision, and recall percent are 96, 97.7, and 97.7, respectively. The comparison unveils the superiority of the suggested approach. Thus, the proposed method effectively improves the performance against the earlier ML techniques and recent DL approaches and avoids overfitting.

Keywords: data stream classification, deep learning, diabetes detection, ensemble technique, extra tree ensemble, machine learning, overfitting, feature selection

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Neha Agrawal,
IIIT Sri City, India
Praveen Kumar,
VIT University, India
Gauri Shankar,
Indian Institute of Technology Dhanbad, India

***Correspondence:**
Hanumat Sastry G
hsastry@ddn.upes.ac.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 October 2021*
Accepted: *16 November 2021*
Published: *15 February 2022*

Citation:
Arya M, Sastry G H, Motwani A, Kumar S and Zaguia A (2022) A Novel Extra Tree Ensemble Optimized DL Framework (ETEODL) for Early Detection of Diabetes. Front. Public Health 9:797877. doi: 10.3389/fpubh.2021.797877

INTRODUCTION

In a survey made by IDF and WHO, nearly half a billion of the population worldwide have diabetes and posing about $13,700 financial burden per year. Moreover, the statistics increasing in years to come. Diabetes is a persistent disease caused when the blood sugar level crosses certain levels and can have adverse consequences on other organs of the human body and severely affect the entire body. If it is not diagnosed at the right time and remains untreated, it can increase the risk of other

disorders. In addition, diabetes may lead to other problems like kidney failure, weak eyesight and heart problem.

Furthermore, diabetic persons are at higher risk of infection and death from COVID-19 (1). Hence, concerning its severe complications, prior diagnosis of the disease is very significant to take timely steps to avoid other health risks and complications. Patients with diabetes can benefit from the Internet of Things devices such as CGM, intelligent pens, and other similar devices, according to the American Diabetes Association (ADA). These devices help in collecting medical data in real-time. Data collected from the smart devices needs to be processed immediately for dissemination of information to practitioners to provide prompt medical attention. Health care data comes from various sources, including medical history and records of patients in hospitals, medical diagnosis reports, medical examination reports by doctors, real-time data from multiple IoT devices and health-related Apps, and data streams from social networking sites. Dealing with such heterogeneous healthcare data has become increasingly difficult in recent years, owing to the large volume of data, security issues, incompetence in wireless network application development, and the rapidity with which it is being generated. As a result, to improve the healthcare industry's efficiency, accuracy, and workflow, data analytics tools are required to manage such complex data (2). Therefore, it is necessary to process this streaming data faster to obtain relevant and significant knowledge using an adaptive model. In the recent past, many algorithms have been employed in handling the classification of real-time data streams. Gaining insights and knowledge from such streaming data is a crucial challenge. Machine Learning (ML) is an analytical method for mechanized learning. It learns from knowledge and skill gained during the training process and applies the experience to improve the performance to make more accurate predictions. ML plays a significant role in discovering hidden and new patterns in the medical data stream. These patterns provide fascinating insights into the knowledge gained.

Further, researchers and medical practitioners use it for assistance in various ailment diagnoses and treatments. Latest advances in this field can also be applied for discovering unknown and latent patterns for detecting diabetes in prior stages. The advancement in ML techniques resolves this censorious problem of an early diabetes diagnosis.

Only health professionals are permitted to process health data because it is sensitive and not readily available. In addition, it is subject to strict usage rules due to the obligation of medical secrecy. As a result, only health professionals are permitted to process it. ML algorithms frequently underperform in prediction accuracy when there is insufficient data to train the model. The data obtained from heterogeneous sources can be either structured or unstructured, depending on the source. Conventional ML techniques cannot process unstructured data; on the other hand, DL can analyze images, videos, and unstructured data in ways that traditional ML techniques cannot. Deep understanding, as opposed to ML, typically necessitates less ongoing human intervention. DL models for prediction and classification are becoming increasingly popular as a means of avoiding these pitfalls.

DL is a subset of Artificial Intelligence with similar architecture as a neural network but has extra hidden layers. The extra layers, therefore, make DL more powerful in data processing than shallow architecture. DL methods showed more accuracy in results than traditional rule-based methods in various domains, including eHealth systems (3–5). A study proposed a hypothesis that further improving the accuracy of DNN feature selection techniques can be used (6, 7). Feature selection is the process of acquiring relevant information and discarding irrelevant ones (8). The feature selection methods can be either supervised or unsupervised, and the supervised method can be divided into the wrapper, filter or intrinsic methods.

Furthermore, a single feature selection method may produce a local optimal or sub-optimal feature subset for which a learning method's performance is compromised. Therefore, multiple feature subsets are combined in the ensemble-based feature selection method to select an optimal subset of features using a combination of feature ranking that improves classification accuracy. The normalized total reduction is the mathematical criteria used in the split decision during the forest construction when the extra tree ensemble method performs feature selection. In addition, in the proposed work Gini Index is computed for each feature known as the Gini Importance of the feature. Then, each feature is ordered in descending order based on its Gini Importance, and the user selects the top k features based on the preferences to perform feature selection. This process selects the optimal feature subset from the high dimensional feature space.

However, the DNN has a disadvantage because it overfits small data sets (9, 10). Overfitting occurs when the accuracy with the training dataset is greater than the accuracy with the testing data set, and the model is not generalized. Dropouts can avoid this problem where a certain number of neurons at a layer are deactivated from firing during training. This deactivation of neurons prevents overfitting, and the network's performance on test data improves (11).

The novelty of the proposed work is as follows:

i. The Extra tree ensemble feature selection technique reduces the feature space by selecting the optimal feature subset. Thus, improving the prediction accuracy and reducing model complexity.
ii. When presented to the DL network, the optimal feature subset further enhances its performance and prevents overfitting.
iii. Previous works are either complex or prone to overfitting. Both these issues are fixed in the proposed work.

Following are the key contributions of the Novel ETEODL framework.

i. The proposed algorithm predicts diabetes at a very early stage using a framework that combines an extra-tree ensemble feature extraction technique used to extract relevant features and a deep neural network to improve prediction accuracy.
ii. Valuable metrics: prediction accuracy, Precision, Recall, F1-score, and computation time are evaluated for performance comparison.
iii. The proposed algorithm is compared with state-of-the-art techniques.

iv. The proposed algorithm shows no instances of overfitting or underfitting.

LITERATURE REVIEW

Alić et al. (12) studied several diverse ML techniques to detect diabetes and concluded that the most common type of ANN used is multilayered feed-forward and the Naïve Bayesian network, which shows higher possibilities of get-ting accurate predictions. In their work, Sisodia and Sisodia (13) experimented with Decision Tree, SVM, and Naive Bayes and concluded that Naive Bayes has the highest accuracy compared to other algorithms and verified it using ROC metrics. Maniruzzaman et al. (14) used Gaussian process classification (GPC) to diagnose diabetes and concluded that the model's performance is comparatively better than other models. In their work, Kaur and Kumari (15) analyzed different models for the detection of diabetes. Wei et al. (16) explored the popular techniques to detect diabetes and data pre-processing techniques. Kamble et al. (47) proposed a DL-based Restricted Boltzmann machine approach for detecting diabetes. Swapna et al. (17) used DL-based methods to classify diabetic and HRV signals by extracting dynamic features from HRV data using a combination of LSTM and CNN. However, LSTM has high computational complexity and is prone to overfitting (18). Yahyaoui et al. (19) compared traditional classifiers with DL-based classifiers, Random Forest (RF) shows more accurate results in predicting diabetes than DL and SVM methods. DL (DL) techniques like CNN and RNN improve performance compared with classic designs (20). Naz and Ahuja (21), in their work, concluded that DL approaches perform better for early detection of diabetes as compared to Artificial Neural Network (ANN), Naive Bayes (NB), and Decision Tree (DT). The DL techniques also facilitate the latest trending techniques like Edge AI applications (22). A regularization layer called dropout can be used in fully connected layers of DNN to address the problem of overfitting (23). Rubaiyat et al. (24) used feature selection and used the selected features with traditional ML techniques like Random Forest, Logistic Regression, and MLP neural network classifier. Iwendi et al. (25) proposed a system to improve intrusion detection using a combination of correlation-based feature selection techniques and machine learning ensemble models. Reddy et al. (26) proposed ensemble-based ML algorithms, compared the proposed performance against the individual ML algorithms and concluded their superiority over the unique ML algorithms. Bashir et al. (27) used ensemble techniques for diabetes detection and inferred that the Bagging ensemble outperforms other ensemble techniques.

Similarly, Tama and Rhee (28) concluded that a tree-based classifier is better than other approaches. Recent research has been concentrated on enhancing the performance of ensemble-based methods for the prediction of disease. In addition, the neural network-based models can further reduce the cost (29). Deepa et al. (30) proposed RASGD to improve the regularization of the classification model. It is done by employing weight decay methods, such as the least absolute shrinkage and selection operator. In addition, ridge regression methods are used to achieve better regularization. Gadekallu et al. (31) were motivated by the fact that previous work had neglected the aspects of data pre-processing and dimensionality reduction, which had resulted in skewed results. Consequently, in their work, the raw dataset is normalized using the Standardscalar technique and Principal Component Analysis (PCA) is used to extract the most significant features from the dataset. In addition, the Firefly algorithm is used to reduce the dimensionality of the data. Finally, this condensed dataset is fed into a Deep Neural Network Model, used to classify the data. Gadekallu et al. (32), in their work, use a principal component analysis-based deep neural network model with the Gray Wolf Optimization (GWO) algorithm. The application of GWO allows for the selection of the most optimal parameters for training the DNN model.

RESEARCH MOTIVATION

In this era of big data, various types of biomedical data for early detection of diabetes have been emerging from electronically generated health records, medical images, IoT sensor data, and simple text data. This streaming data is intricate, diverse, appallingly annotated, and commonly not structured. Early detection of the disease is significant to take timely steps to avoid other health risks and complications. However, the earlier studies depict that the ML algorithms need structured data for classification, have less prediction accuracy, prone to overfitting, and require more computational time to predict the disease. While DL models are more promising, as DL networks are flexible, making them suitable for structured and unstructured data (33), they can process the data than the shallow architecture.

Further, their accuracy can be improved by appropriate feature selection techniques (34, 35). Thus, DL approaches could employ big biomedical data to improve human health (36). For example, most of the recent research (20) concluded that DL performs better for early detection of diabetes as compared to Artificial Neural Network (ANN), Naive Bayes (NB), and Decision Tree (DT).

In literature, various techniques like linear regression feature selection, logistic regression feature selection, Correlation-value based feature selection, Chi- square-based feature selection, F-score based feature selection, decision tree feature selection and random forest feature selection are used for selecting relevant features.

Table 1 below summarizes the methodology used and the limitations of some of the pertinent recent works.

Chen et al. (42) concluded that feature extraction might improve the performance of deep neural networks. In their work, Motwani et al. (43) suggested a framework based on a deep neural network for intelligent patient monitoring. Motwani et al. (44) used DL with cost optimization for remote patient monitoring and recommendation. Authors in (45) compared various feature selection techniques and concluded that the random forest algorithm performs better than the other algorithms. Trapping this advantage of the tree-based technique for selecting relevant features based on feature importance, the proposed approach uses the Extra tree ensemble feature selection technique to

TABLE 1 | Methodology and limitations of recent relevant work.

S.No	Author	Methodology used	Limitations
1	Cho et al. (37)	A model which combines Linear SVM classifiers and wrapper or embedded feature selection methods	Wrapper methods for feature selection have high computational costs and are generally prone to overfitting. They are also dependent on the classifiers used.
2	Le et al. (38)	A novel model utilizing Gray Wolf Optimization (GWO) and an Adaptive Particle Swam Optimization (APSO) to optimize the Multilayer Perceptron (MLP) to reduce the number of required input attributes.	In MLP, computations are complex and time-consuming.
3	Lukmanto et al. (39)	A classification framework to identify and classify diabetes datasets using F-Score Feature Selection and Fuzzy SVM.	A disadvantage of the F-score is that it does not reveal mutual information among features. Instead, it only captures the linear relationships between features and labels.
4	Putri et al. (40)	Learning Vector Quantization (LVQ) to classify the diabetes dataset with Chi-Square for feature selection.	Chi-Square for feature selection does not take into consideration the feature interactions. It is best suited only for categorical variables
5	Sneha and Gangil (41)	Classification by selecting the optimal features based on the correlation values.	Correlation values for feature selection uncover only relationships and do not determine what variables have the most influence. Thus, it can be a time-consuming process.

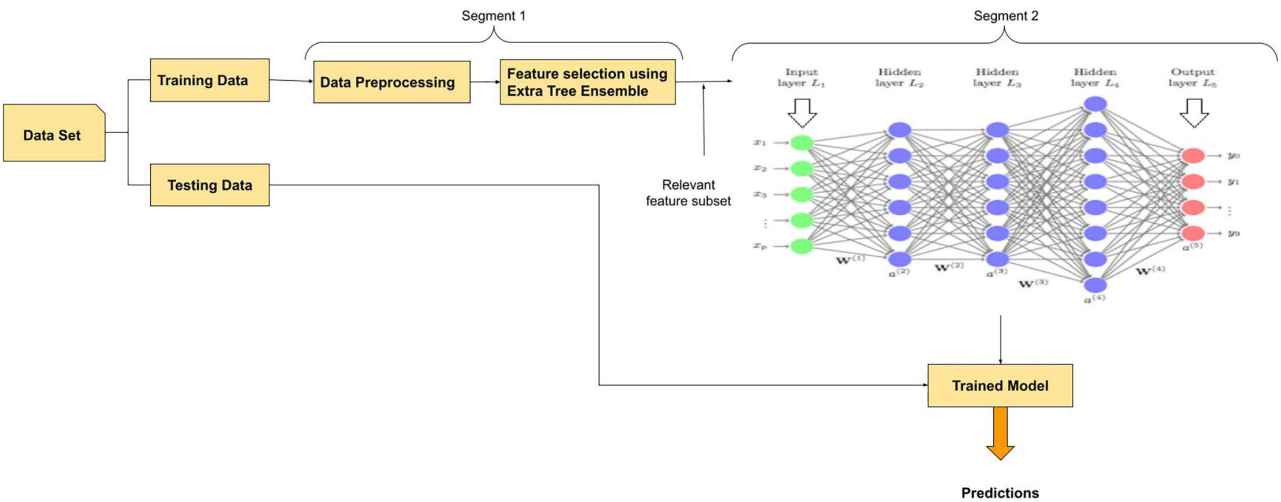


FIGURE 1 | Flowchart for the proposed framework.

retrieve the optimal feature set. Furthermore, a single feature selection method may produce a local optimal or sub-optimal feature subset for which a learning method's performance is compromised. Therefore, multiple feature subsets are combined in the ensemble-based feature selection method to select an optimal subset of features using a combination of feature ranking that improves classification accuracy.

Thrust by these facts, this paper proposed an Extra-Tree Ensemble optimized DL framework (ETEODL) to predict the likelihood of diabetes. This approach is a combination DL approach for prediction and an Extra Tree ensemble technique for selecting the best features based on feature importance. The DL approach extracts lower-level information and feeds them to the next higher layer. The dropout technique is

TABLE 2 | Comparison of proposed model with conventional ML algorithms.

Classifier	Accuracy%	Precision %	Recall %	F1-score %	ROC area %	Comp time
Naïve bayes	87.5	88.2	87.5	87.6	94	38.65
Decision tree	80.76	85.3	80.7	81.1	83.7	45.56
Hoeffding tree	87.5	88.2	87.5	87.6	94	50.68
Random forest	95.19	95.55	95.19	95.2	91.1	54.72
Ensemble (stacking)	63.46	83.4	63.46	73.5	50	69.8
ETEODL (proposed)	97.38	97.7	97.7	96	95	28.63

TABLE 3 | Comparison of the proposed model with recent works.

Title	Methodology	Accuracy	Precision	Recall	F1-score	ROC area	Computation time	Limitation/ drawback
Diabetes detection using DL algorithms (17)	Employed long short-term memory (LSTM), convolutional neural network (CNN), and its combinations for extracting complex temporal dynamic features	95.7	0.77	0.86	0.87	0.94	58.73	LSTM have high computational complexity and is prone to overfitting
Health care system: stream ML classifier for features prediction in diabetes therapy (46)	Used combination of probabilistic and ML models	90	0.746	0.678	0.85	0.5	45.67	The Probabilistic approach suffers from the problem of selecting the suitable metrics to conduct a detection process
Diabetes detection using DL approach (47)	DL-based Restricted Boltzmann machine approach is used.	84.32	0.86	0.75	0.77	0.911	67.83	In RBM, training is more problematic as it is difficult to calculate the energy gradient function
ETEODL (Proposed)		97.38	0.977	0.977	0.96	0.95	28.63	

used with hidden layers of DNN to prevent the overfitting of the model.

The proposed model performs better in the following aspects:

(1) Prediction accuracy and computational time
(2) Prevent overfitting.

The framework is compared with traditional ML techniques and recent works for various parameters like Accuracy, F1-Score, Precision, Recall, and Computation time and thus concluded that the efficiency of the proposed framework is better than the compared works.

The paper has the following sections. First, in section Methodology and Algorithm of the proposed framework are discussed. Then, in section 5, Experimental Setup, including Dataset description, Experiment Environment setup, Experiments, and results, has been discussed. In the last, the proposed work is concluded, and directions for future research are suggested.

METHODOLOGY AND ALGORITHM

The DL framework used in this paper for the early prediction of diabetes is called ETEODL. The framework is divided into two segments. The first segment performs the data acquisition, pre-processing, and feature extraction, while the second segment performs the prediction using a DL model. First, data is acquired from the UCI repository. The acquired data is pre-processed to make it ready for further processing. Next, the data set is split for training and testing purposes. It is followed by feature extraction, where relevant features are extracted to reduce the data set's feature space and prevent model

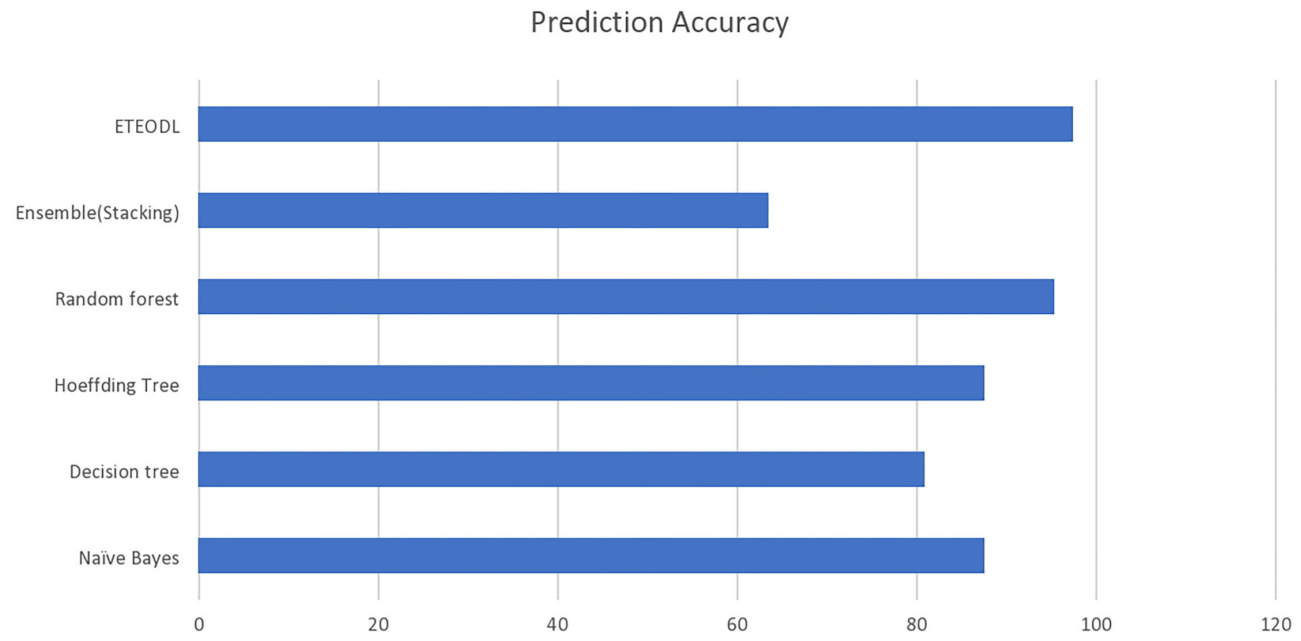


FIGURE 2 | Accuracy (%) comparison with existing methods.

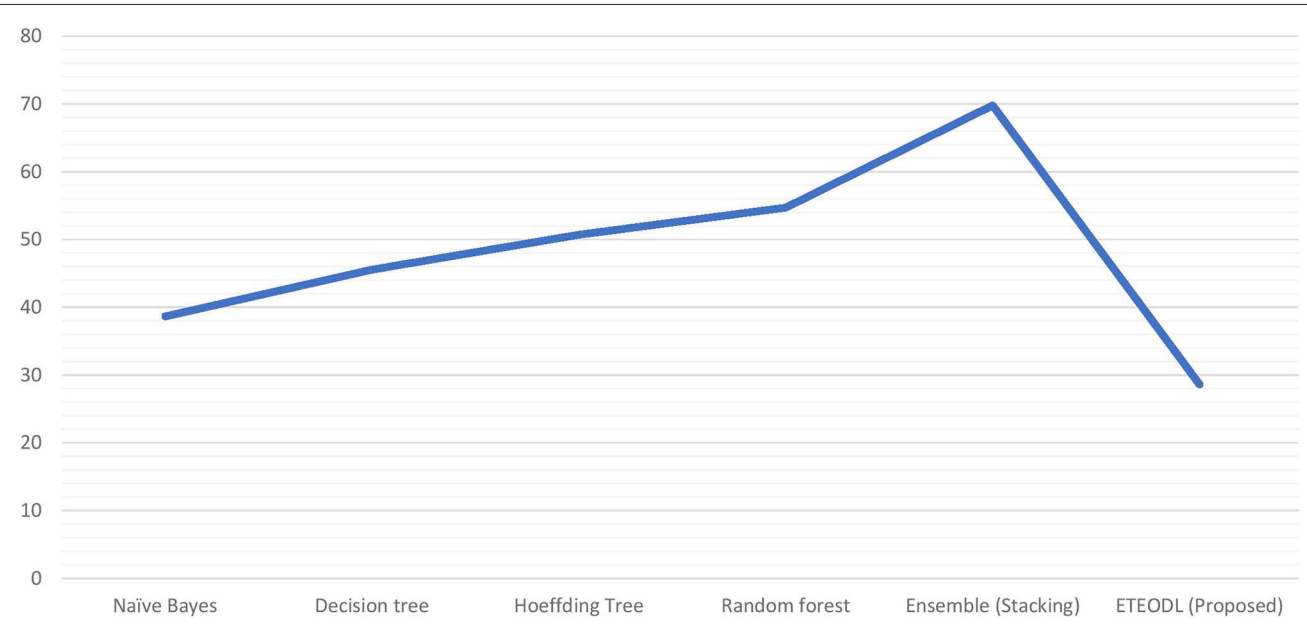


FIGURE 3 | Comparison of computation time (in sec).

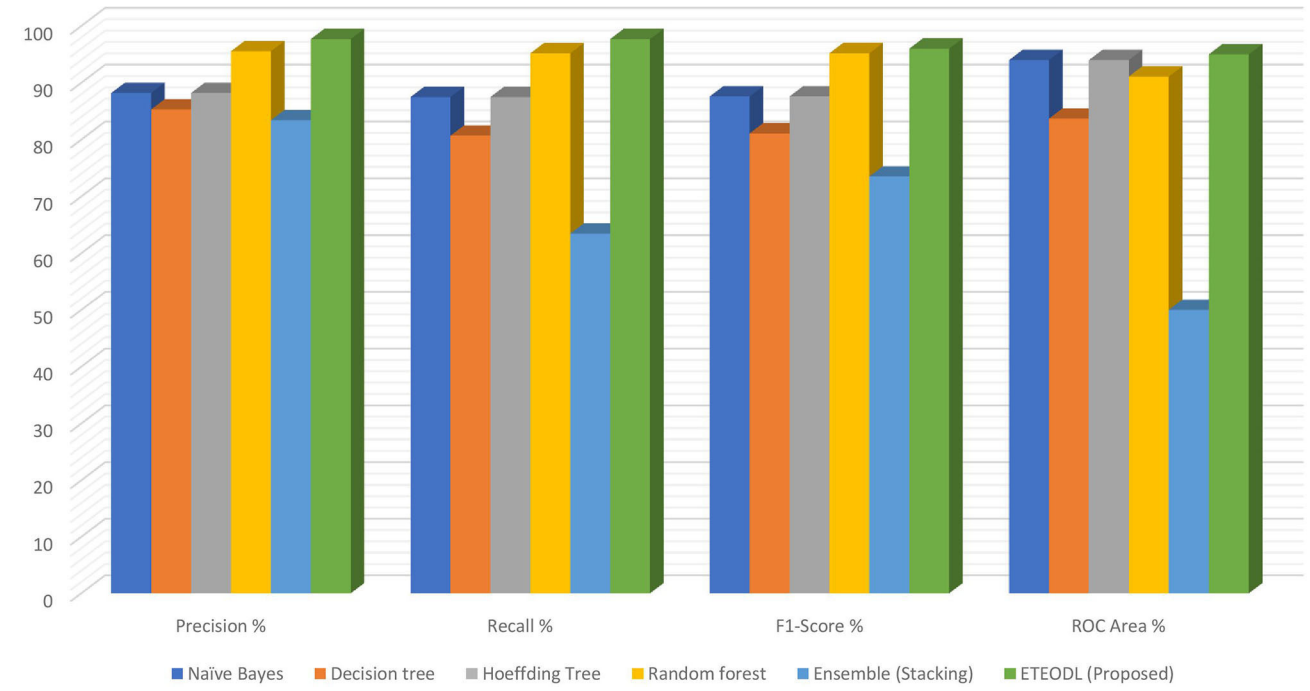


FIGURE 4 | Comparison of precision, recall, F1-score, and ROC area.

overfitting. The extra tree feature extraction technique is used in the proposed work. The output of the first segment is fed in a DL model for classification and prediction. The functions of the two segments are explained in detail in the following section.

Segment-1

Data Acquisition and Data Pre-processing

Data is acquired and pre-processed in the first phase of the framework to clean, transform, and reduce the dimensions. Then, the missing values are discarded. Finally, the normalization process does data transformation. The dataset is divided in a ratio of 80:20 for training and testing purposes for ensuring the learning process only from the training data. After training the model, the performance is tested using testing data.

Data Feature Selection Using Extra Tree Ensemble

After data pre-processing, feature selection is made using Extra Tree Ensemble technique. In this step, the subset of the most relevant features is selected. The choice of the most pertinent features influences the model performance greatly. In the proposed framework, feature importance property is utilized for feature selection. In this method, each feature is given a score. The score ranges from zero and one. The leading score indicates more relevancy of the feature toward the targeted output. These relevant features are thus chosen for model building with improved predictive accuracy and controlled overfitting. A subset of randomly selected n features is supplied to each test node of the tree. Further, the best feature is chosen by the decision tree from this subset to split the data based on Gini Index. The output features of segment one are used as input in segment 2.

Segment-2

DL Model With Dropouts

The DL model consists of three types of layers:

The Input layer is where the selected features are passed. No computation occurs in this layer. The Hidden layers are present between the input and output layers. For choosing the number of Hidden Layers following basic rules are followed;

(1) If the data is linearly separable, then no hidden layers are required.
(2) Using neural networks with one to two hidden layers would be appropriate if the data is less complex and has fewer dimensions or features.
(3) If the data has many dimensions or features, it is possible to use three to five hidden layers to achieve the best possible result.

The input data is not linearly separable and is complex as obtained from various heterogeneous sources; the proposed model consists of three hidden layers. The hidden layers one and two utilize the Rectified Linear Unit (ReLU) as activation.

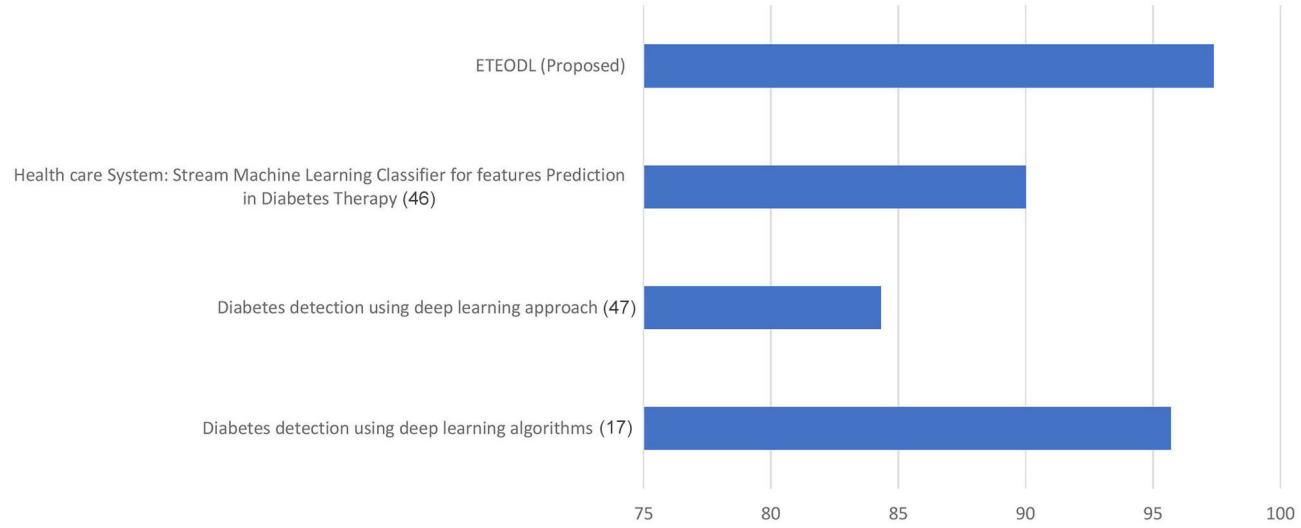


FIGURE 5 | Comparison of prediction accuracy.

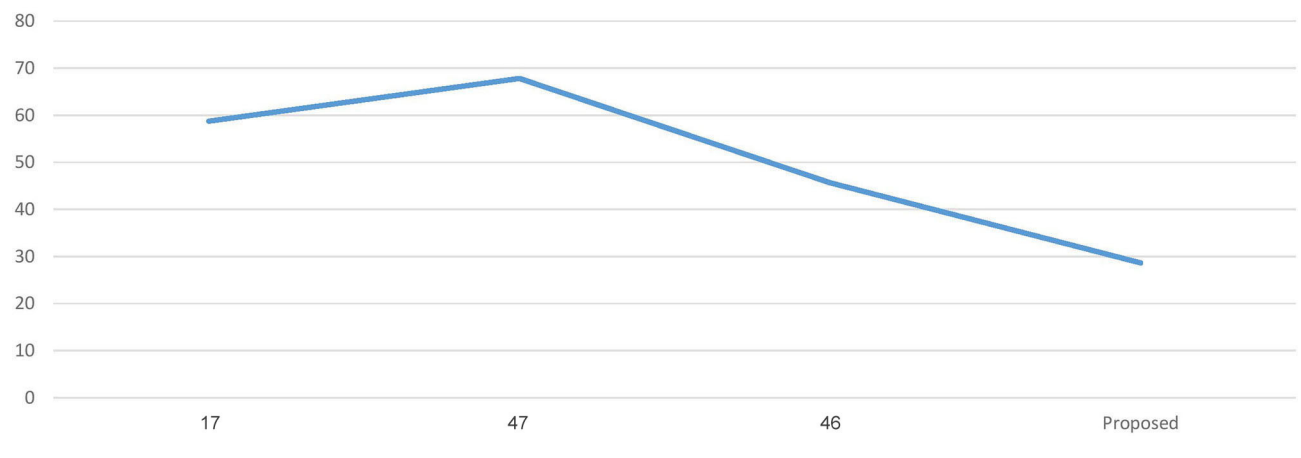


FIGURE 6 | Comparison of computation time.

Mathematically, it is defined as

$$y = max(0, x) \tag{1}$$

And at hidden layer three uses the sigmoid as an activation function. The sigmoid is mathematically represented as:

$$f(x) = 1/(1 + e^{\wedge}(-x)) \tag{2}$$

These hidden layers perform the computation and pass the information to the output layer in the end.

The output layer is responsible for producing the out variable and giving the result.

The cost or loss function is binary cross-entropy/log loss represented using E(W).

$$[E(W)] = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(\gamma) + (1 - y_i)\log(1 - \gamma) \tag{3}$$

A dropout layer is added to the model after each hidden layer to prevent overfitting. During the training phase, the dropout layer deactivates a random set of fractions "i" neurons. The value of p is set to 0.8, where p is the probability of retention used in the input layers and is set to 0.5 in the hidden layers. The value of c is set to 4 in all the layers, where c is the Max-norm constraint. The step-wise methodology of the framework is shown in **Figure 1**.

The algorithm of the proposed framework is given in below.

Algorithm 1| Extra Tree Ensemble Optimized DL Algorithm

Input: X: Training Data set $\tilde{y}$
Y: Class Label of *X*
x: Unknown sample
Output: Label k for unseen sample x
1: Call Algorithm-2 for ETE on Dataset X;
2: Ensemble $x_n = h(\hat{y}_1, \hat{y}_2, \ldots\ldots \hat{y}_n)$
3: Transform Input Features: x_n *to Tensor* Tx_n
4: for each hidden layer, l do

a. $r_j\,(\,l\,) \sim$ *Bernaulli* (p)
b. $\tilde{y}(\,l\,) = r\,(\,l\,)\,{}^{*}y\,(\,l\,)$
c. $z_i\ (\,l+1\,) = w_i\,(\,l+1\,)\,{}^{*}\tilde{y}\,(\,l\,) + b_i\ (\,l+1)$
d. $y_i\,(\,l+1\,) = f\,(z_i\,(\,l+1\,))$

5: Calculate the probability score for predicting the class of transaction:

$$\hat{y} = \sigma(\,Wo\,[Td_m\,] + b_f\,)$$

6: Calculate objective function such as Error Function: E(W) is calculated as

$$[E\,(W)] = -\frac{1}{m}\sum_{i=1}^{m} y_i \log(\gamma) + (1-y_i)\log(1-\gamma)$$

7: Predict diabetes for the given feature set.

Algorithm 2| ETE

Input: F
Output: A Split [f < fc] or none
1: if ($|F| < n_{\min}$) then
return SS=True;
2: if all features are constant in F
return SS=True;
3: if the output is constant in F
return SS=True;
else
return SS=false;
Endif
4: if (SS) is TRUE
return none;
else

a) Select k features $\{f_1, f_2, \ldots\ldots\ldots\ldots f_k\}$ from all f_c that is variable (in F);
b) Take k splits $\{s_1, s_2, \ldots\ldots s_k\}$ where s_i = call Algorithm-3 Rand_Split (s, f_i) $\forall_{\ i}$ =1,2,...k;
c) return a split s* such that Score (s*, F) = max i=1 to k Score (s*, F);

End if

Algorithm 3| Rand_Split(s,f)

Input: s and f
Output: a split
1: Draw a random cut point f_k uniformly in $[fs_{\min}, fs_{\max}]$;
2: return the split $[a < a_c]$;

EXPERIMENTAL SETUP AND RESULTS

Data Set

The dataset used in the research is taken from the UCI repository (34). It contains 520 instances and 16 attributes. The missing values have been pre-processed by discarding the tuples with incomplete values.

Experiment Environment Setup

The experimental setup includes an Intel Core i5 processor with 16 GB RAM. The software configuration includes Keras, Google Tensorflow, and other required libraries such as Scikit-Learn, Numpy, and Pandas installed over Python.

Results and Discussion

The proposed ETEODL model was implemented over Python. The parameters to evaluate the model. The model efficiency was evaluated based on essential metrics like prediction accuracy, precision, recall, f1-measure, ROC, and RMS Error (14). The proposed model is compared with conventional techniques (13) like Naïve Bayes, Decision Tree, Random-forest, Hoeffding tree, and ensemble classifier-like stacking and also with the recent three related works (17, 18, 46). The comparisons are with conventional work is summarized in **Table 2**, and comparison with recent existing work is summarized in **Table 3**.

Table 2 shows the comparison of the proposed framework and conventional ML algorithms.

The outcome of the comparison in **Table 2** can be concluded as follows:

i The DL-based framework outperforms traditional algorithms for the early detection of diabetes (21).
ii The Extra-Tree ensemble feature extraction technique prevents the overfitting of the DL model (28).

The graph in **Figure 2** compares the prediction accuracy of conventional ML algorithms with the ETEODL (proposed).

Figure 3 represents the graphs comparing the computation time of conventional ML algorithms with the ETEODL (proposed).

Figure 4 represents the graphs comparing the precision, recall, and F1 score.

Table 3 shows the performance comparison of the proposed framework with Recent work.

The proposed work is less computationally complex and requires less computation time than previous related work while improving performance.

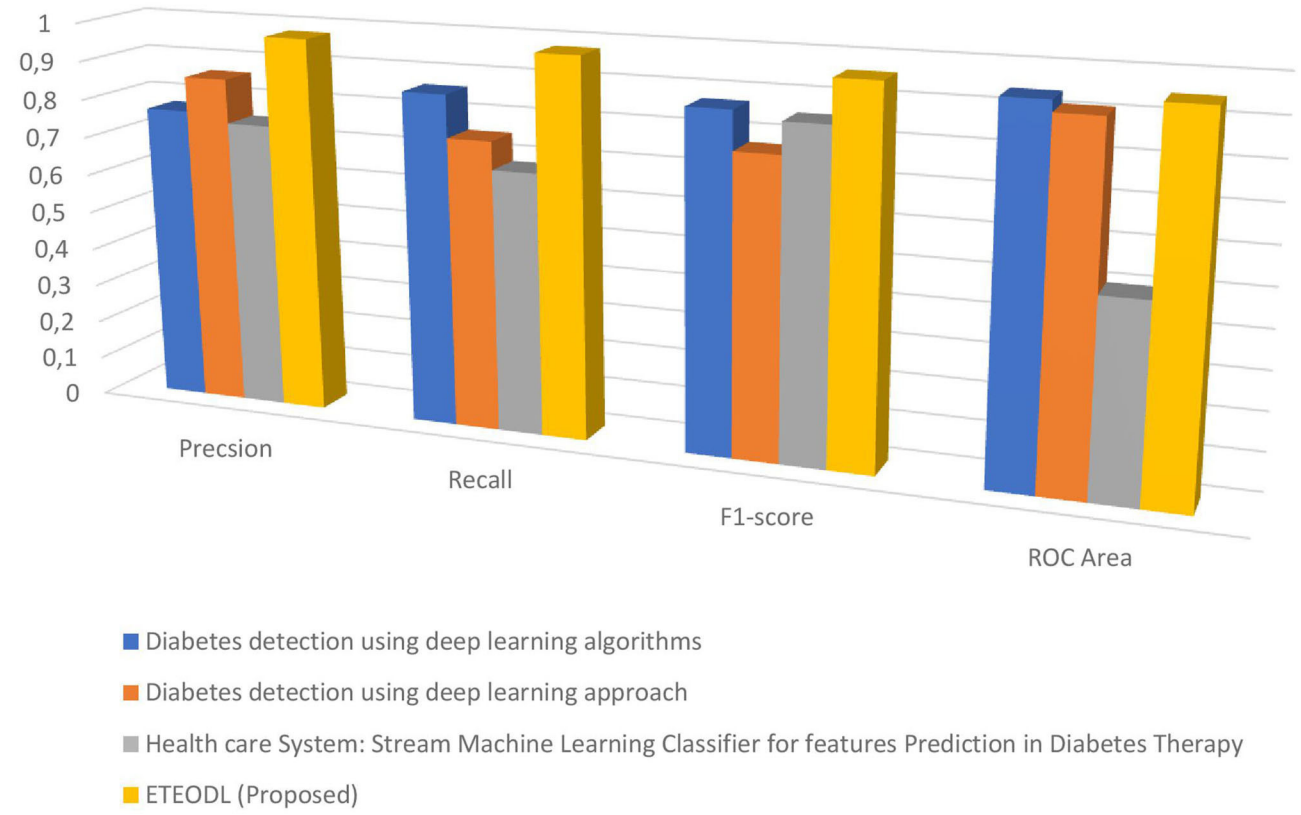


FIGURE 7 | Comparison of precision, recall, f1-score, and roc area.

Figure 5 represents the graph comparing the accuracy of the proposed model with Recent works (17, 46, 47).

Figure 6 represents the graph Comparing the computation time of the Proposed model with three Recent works.

Figure 7 represents the graph Comparing the proposed model's Precision, Recall, and F1 score with Recent works.

Table 4 compares the training and testing accuracy of ETEODL over hundred epochs.

The graph in **Figure 8** shows the comparison of training and testing accuracies.

The comparison graph shows that the proposed work prevents overfitting as it maintains good training and testing accuracies.

TABLE 4 | Comparison of training and testing accuracy of ETEODL over 100 epochs.

EPOCHS	Average training accuracy	Average testing accuracy
01–10	0.836	0.954
11–20	0.877	0.962
21–30	0.922	0.968
31–40	0.934	0.971
41–50	0.952	0.973
51–60	0.964	0.976
61–70	0.965	0.977
71–80	0.966	0.977
81–90	0.968	0.978
91–100	0.968	0.978

CONCLUSION

Detecting diabetes at an early stage is very consequential to take well-timed steps to avoid other health risks and complications. The proposed work demonstrates the application of the Extra Tree ensemble for optimization and DL for classification and prediction for diabetes. The proposed ETEODL is compared against the conventional ML techniques and ensemble algorithms and similar recent works. The accuracy obtained using ETEDOL was approximately 97.38% which is better than contemporary and traditional techniques. The information thus predicted can provide a caution signal for the patient and the doctor to take precautions and control measures. The proposed method is limited to numerical and categorical data and can also be developed

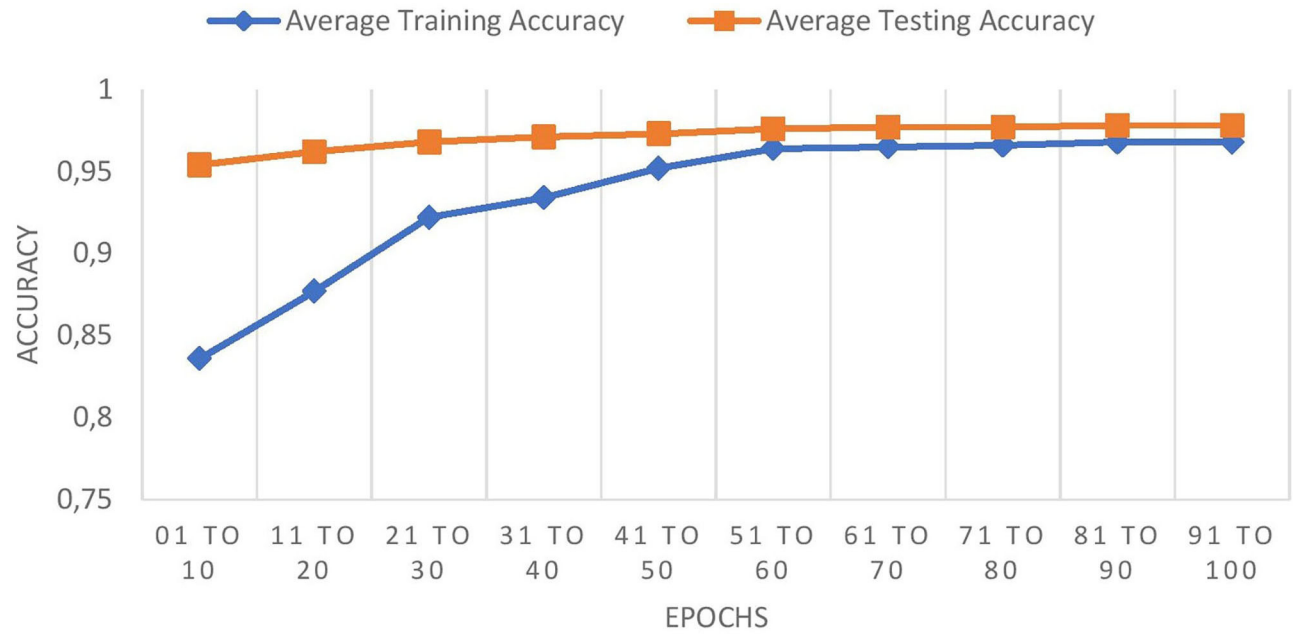


FIGURE 8 | Comparison of training and testing accuracy.

for image data. In the future, the proposed model can be extended as an integral part of an automated system for diabetes prediction.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MA and HS developed and implemented algorithm. AM, SK, and MA evaluated the performance of the algorithm and created the first draft. AZ further evaluated and benchmarked the work. HS and AZ finalized the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Taif University Researchers Supporting Project Number (TURSP-2020/114), Taif University, Taif, Saudi Arabia.

ACKNOWLEDGMENTS

This work was supported by Taif University Researchers Supporting Project Number (TURSP-2020/114), Taif University, Taif, Saudi Arabia. Authors express thanks to University of Petroleum and Energy Studies, India, Taif Univerisity, Saudi Arabia and Bhilai Institute of Technology, India for Providing the state of art research facilities.

REFERENCES

1. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* (2019) 19:281. doi: 10.1186/s12911-019-1004-8
2. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
3. Stephen O, Sain M, Maduh UJ, Jeong DU. An efficient deep learning approach to pneumonia classification in healthcare. *J Healthc Eng.* (2019) 2019:4180949. doi: 10.1155/2019/4180949
4. Tomov NS, Tomov S. *On Deep Neural Networks for Detecting Heart Disease*. (2018). Available online at: http://arxiv.org/abs/1808.07168 (accessed November 11, 2021).

5. Kutia S, Chauhdary SH, Iwendi C, Liu L, Wang Y, Bashira AK. Socio-technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access*. (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
6. Liu Y, Yu Z, Sun H. Prediction method of gestational diabetes based on electronic medical record data. *J Healthc Eng*. (2021) 2021:6672072. doi: 10.1155/2021/6672072
7. Huang M, Huang C, Yuan J, Kong D. A semiautomated deep learning approach for pancreas segmentation. *J Healthc Eng*. (2021):3284493. doi: 10.1155/2021/3284493
8. Bourouis S, Alroobaea R, Rubaiee S, Andejany M, Bouguila N. Nonparametric bayesian learning of infinite multivariate generalized normal mixture models and its applications. *Appl Sci*. (2021) 11:5798. doi: 10.3390/app11135798
9. Cogswell M, Ahmed F, Girshick R, Zitnick L, Batra D. Reducing overfitting in deep networks by decorrelating representations. *arXiv preprint arXiv:1511.06068*. (2015).
10. Salman S, Liu X. *Overfitting Mechanism and Avoidance in Deep Neural Networks*. (2019). Available online at: http://arxiv.org/abs/1901.06566 (accessed November 11, 2021).
11. Mele B, Altarelli G. Lepton spectra as a measure of b quark polarization at LEP. *Phys Lett B*. (1993) 299:345–50. doi: 10.1016/0370-2693(93)90272-J
12. Alić B, Gurbeta L, Badnjević A. Machine learning techniques for classification of diabetes and cardiovascular diseases. In: *2017 6th Mediterranean Conference on Embedded Computing (MECO)*. IEEE (2017). p. 1–4.
13. Sisodia D, Sisodia DS. Prediction of diabetes using classification algorithms. *Procedia Comput. Sci*. (2018) 132:1578–85. doi: 10.1016/j.procs.2018.05.122
14. Maniruzzaman M, Kumar N, Menhazul Abedin M, Shaykhul Islam M, Suri HS, El-Baz AS, et al. Comparative approaches for classification of diabetes mellitus data: machine learning paradigm. *Comput Methods Programs Biomed*. (2017) 152:23–34. doi: 10.1016/j.cmpb.2017.09.004
15. Kaur H, Kumari V. Predictive modelling and analytics for diabetes using a machine learning approach. *Appl Comput Informatics*. (2020). doi: 10.1016/j.aci.2018.12.004. [Epub ahead of print].
16. Wei S, Zhao X, Miao C. A comprehensive exploration to the machine learning techniques for diabetes identification. In; *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE (2018). p. 291–5.
17. Swapna G, Vinayakumar R, Soman KP. Diabetes detection using deep learning algorithms. *ICT Express*. (2018) 4:243–6. doi: 10.1016/j.icte.2018.10.005
18. Duc TN, Minh CT, Xuan TP, Kamioka E. Convolutional neural networks for continuous QoE prediction in video streaming services. *IEEE Access*. (2020) 8:116268–78. doi: 10.1109/ACCESS.2020.3004125
19. Yahyaoui, Jamil A, Rasheed J, Yesiltepe M. A decision support system for diabetes prediction using machine learning and deep learning techniques. In; *2019 1st International Informatics and Software Engineering Conference (UBMYK)*. IEEE (2019). p. 1–4.
20. Almulihi S, Alharithi AH, Mechti FS, Alroobaea S, Rubaiee R. A software for thorax images analysis based on deep learning. *Int J Open Source Softw Process*. (2021) 12:60–71. doi: 10.4018/IJOSSP.2021010104
21. Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J Diabetes Metab Disord*. (2020) 19:391–403. doi: 10.1007/s40200-020-00520-5
22. Masud M, Singh P, Gaba GS, Kaur A, Alghamdi RA, Alrashoud M et al. CROWD: crow search and deep learning based feature extractor for classification of parkinson's disease. *ACM Tran Internet Technol*. (2021) 21:1–18. doi: 10.1145/3418500
23. Ashiquzzaman A, Tushar AK, Islam MR, Shon S, Im K, Park JH, et al. Reduction of overfitting in diabetes prediction using deep learning neural network. *Lect Notes Electr Eng*. (2017) 449:35–43. doi: 10.1007/978-981-10-6451-7_5
24. Rubaiat SY, Rahman MM, Hasan MK. Important feature selection accuracy comparisons of different machine learning models for early diabetes detection. In; *2018 International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE (2018). p. 1–6.
25. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M. The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors*. (2020) 20:1–37. doi: 10.3390/s20092559
26. Reddy GT, BhattacharyaS, Sivaramakrishnan S, Chowdhary CL, Hakak S, Kaluri R, et al. An ensemble based machine learning model for diabetic retinopathy classification. In; *Int Conf Emerg Trends Inf Techno. Eng. ic-ETITE 2020* (2020).
27. Bashir S, Qamar U, Khan FH, Javed MY. An efficient rule-based classification of diabetes using ID3, C4.5, & CART ensembles. In; *2014 12th International Conference on Frontiers of Information Technology*. IEEE (2014). p. 226–31.
28. Tama BA, Rhee KH. Tree-based classifier ensembles for early detection method of diabetes: an exploratory study. *Artif. Intell. Rev*. (2019) 51:355–70. doi: 10.1007/s10462-017-9565-3
29. Liu H, Yue K, Cheng S, Pan C, Sun J, Li W. Hybrid model structure for diabetic retinopathy classification. *J Healthc Eng*. (2020) 2020:8840174. doi: 10.1155/2020/8840174
30. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-adaline stochastic gradient descent classifier. *J Supercomput*. (2021) 77:1998–2017. doi: 10.1007/s11227-020-03347-2
31. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Reddy P, Ra IH, et al. Early detection of diabetic retinopathy using pca-firefly based deep learning model. *Electron*. (2020) 9:274. doi: 10.3390/electronics9020274
32. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intell Humaniz Comput*. (2020) 24:1–4. doi: 10.1007/s12652-020-01963-7
33. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models : a deep learning approach. *BMC Med Inform Decis Mak*. (2020) 20:280. doi: 10.1186/s12911-020-01297-6
34. Gupta M, Konar D, Bhattacharyya S, Biswas S. *Computer Vision and Machine Intelligence in Medical Image Analysis*. Singapore: Springer (2020).
35. Rathi T. *Variable Weights Neural Network For Diabetes Classification*. arXiv preprint arXiv:2102.12984. (2021).
36. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare : review , opportunities and challenges. *Brief Bioinform*. (2018) 19:1236–46. doi: 10.1093/bib/bbx044
37. Cho BH, Yu H, Kim KW, Kim TH, Kim IY, Kim SI. Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artif Intell Med*. (2008) 42:37–53. doi: 10.1016/j.artmed.2007.09.005
38. Le TM, Vo TM, Pham TM, Svt DAO. A novel wrapper — based feature selection for early diabetes prediction enhanced with a metaheuristic. *IEEE Access*. (2021) 9:7869–84. doi: 10.1109/ACCESS.2020.3047942
39. Lukmanto RB, Suharjito D, Nugroho A, Akbar H. Sciencedirect sciencedirect early detection of diabetes mellitus using feature selection and fuzzy support vector machine. *Procedia Comput Sci*. (2019) 157:46–54. doi: 10.1016/j.procs.2019.08.140
40. Putri NK, Rustam Z, Sarwinda D. Learning vector quantization for diabetes data classification with chi-square feature selection. *IOP Conference Series: Mat Sci Eng*. (2019) 546:052059.doi: 10.1088/1757-899X/546/5/052059

41. Sneha N, Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. *J Big Data*. (2019) 6:13. doi: 10.1186/s40537-019-0175-6
42. Chen Z, Pang M, Zhao Z, Li S, MiaoR, Zhang Y, et al. Data and text mining feature selection may improve deep neural networks for the bioinformatics problems. *Bioinformatics*. (2020) 36:1542–52. doi: 10.1093/bioinformatics/btz763
43. Motwani A, Shukla PK, Pawar M. Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation (SPMR). *J Ambient Intell Humaniz Comput*. (2021) 2:1–6. doi: 10.1007/s12652-020-02790-6
44. Motwani M, Shukla A, Pawar PK. Smart predictive healthcare framework for remote patient monitoring and recommendation using DL with novel cost optimization. In; *International Conference on Information and Communication Technology for Intelligent Systems*. Singapore: Springer (2020).
45. Oladimeji OO. Classification models for likelihood prediction of diabetes at early stage using feature selection. *Appl Comput Inform*. (2021). doi: 10.1108/ACI-01-2021-0022. [Epub ahead of print].
46. Ramana D. Health care system : stream machine learning classifier for features prediction in diabetes therapy. *Int J Appl Eng Res*. (2018) 13:59–65. Available online at: https://www.ripublication.com/ijaer18/ijaerv13n1_09.pdf
47. Kamble ST, Patil MTP. Diabetes detection using deep learning approach. *Int J Innov Res Sci Technol*. (2016) 2:342–9.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer PK has declared a shared parent affiliation with the author AM at the time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Arya, Sastry G, Motwani, Kumar and Zaguia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 22 February 2022
doi: 10.3389/fpubh.2022.840525

Development and Validation of a Competitive Risk Model in Elderly Patients With Chromophobe Cell Renal Carcinoma: A Population-Based Study

Jinkui Wang[1,2,3,4,5,6,7†], *Chenghao Zhanghuang*[1,2,3,4,5,6,7,8,9†], *Xiaojun Tan*[1,2,3,4,5,6,7], *Tao Mi*[1,2,3,4,5,6,7], *Jiayan Liu*[1,2,3,4,5,6,7], *Liming Jin*[1,2,3,4,5,6,7], *Mujie Li*[1,2,3,4,5,6,7], *Zhaoxia Zhang*[1,2,3,4,5,6,7] and *Dawei He*[1,2,3,4,5,6,7*]

[1] Department of Urology, Children's Hospital of Chongqing Medical University, Chongqing, China, [2] Chongqing Key Laboratory of Children Urogenital Development and Tissue Engineering, Chongqing, China, [3] Chongqing Key Laboratory of Pediatrics, Chongqing, China, [4] Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China, [5] National Clinical Research Center for Child Health and Disorders, Chongqing, China, [6] China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Chongqing, China, [7] Children's Hospital of Chongqing Medical University, Chongqing, China, [8] Department of Urology, Kunming Children's Hospital, Kunming, China, [9] Yunnan Provincial Key Research Laboratory of Pediatric Major Diseases, Kunming, China

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Praveen Kumar,
VIT University, India
Ebuka Ibeke,
Robert Gordon University, United Kingdom

***Correspondence:**
Dawei He
hedawei@hospital.cqmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 21 December 2021
Accepted: 11 January 2022
Published: 22 February 2022

Citation:
Wang J, Zhanghuang C, Tan X, Mi T, Liu J, Jin L, Li M, Zhang Z and He D (2022) Development and Validation of a Competitive Risk Model in Elderly Patients With Chromophobe Cell Renal Carcinoma: A Population-Based Study. Front. Public Health 10:840525. doi: 10.3389/fpubh.2022.840525

Background: Renal cell carcinoma (RCC) is the most common renal malignancy in adults, and chromophobe renal cell carcinoma (chRCC) is the third most common subtype of RCC. We aimed to construct a competitive risk model to predict cancer-specific survival (CSS) in elderly patients with chRCC.

Methods: The clinicopathological information of the patients was downloaded from the SEER database, and the patients were randomly divided into the training and validation cohorts. Patients' risk factors for cancer-specific death (CSM) were analyzed using proportional subdistribution hazard (SH). We constructed a competitive risk model to predict the CSS of elderly chRCC patients. Consistency index (C-index), the area under receiver operating curve (AUC), and a calibration curve were used to validate the model's accuracy. Decision curve analysis (DCA) was used to test the clinical value of the model.

Results: A total of 3,522 elderly patients with chRCC were included in the analysis. Patients were randomly assigned to either the training cohort (N = 2,474) or the validation cohort (N = 1,048). SH analysis found that age, race, T, N, and M stage, tumor size, and surgery were risk factors for CSM. We constructed a competitive risk model to predict patients' CSS. In the training set, the model predicted patients' 1-, 3-, and 5-year CSS with C-indices of 82.2, 80.8, and 78.2, respectively. The model predicted patient 1-, 3-, and 5-year CSS in the validation cohort with C-indices of 84.7, 83.4, and 76.9, respectively. The calibration curve showed that the model's predicted value is almost consistent with the observed value, which indicated that the model has good accuracy. The AUC of the training set and validation queue also suggested that the model has good discrimination. The clinical utility of the DCA model in predicting patients' CSS is higher than that of traditional TNM staging.

Conclusions: We constructed a competitive risk model to predict CSS in elderly patients with chRCC. The model has good accuracy and reliability, which can help doctors and patients to make clinical decisions and follow-up strategies.

Keywords: competitive risk model, nomogram, elderly, chromophobe cell renal carcinoma, SEER

INTRODUCTION

Renal cell carcinoma (RCC) is the most common renal malignancy in adults, with more than 400,000 cases diagnosed each year (1). According to its pathological classification, it is mainly divided into clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) and chromophobe cell renal cell carcinoma (chRCC), etc. In 2016, The World Health Organization (WHO) defined chRCC as the third common subtype of RCC (2). ChRCC accounts for about 5–7% of RCC, with the same proportion in males and females, and the first diagnosis age of most cases is after 65 years old (3). Most patients are sporadic, and a few are hereditary, including BIRt-Hogg-Dube (BHD) syndrome (4) and tuberous sclerosis (TSC) (5). Comprehensive genomic analysis of chRCC showed a low mutation rate in its somatic cells, and the most common mutated genes were identified as TP53 and PTEN (6). ChRCC originates from the epithelial cells of the distal renal tubules, especially the α-intercalated cells of the renal collecting tubules. CcRCC derived from proximal renal tubular epithelial cells, which provided evidence for the heterogenesis of the two. ChRCC has a good prognosis, with a 5-year survival rate of 78–100% and a 10-year survival rate of 80–90% (7), which is significantly better than other RCCs, including ccRCC. So to this day, research on chRCC remains extremely rare. However, it is worth noting that 5% of chRCC patients are complicated with severe renal venous carcinoma thrombi (8), while nearly 10% of patients develop metastasis (9). The prognosis of these patients is similar to that of clear metastatic renal cell carcinoma, and the 5-year survival rate is only 14% (10). In addition, according to previous reports, factors determining the survival of chRCC patients also vary greatly (11, 12). Therefore, it is of great significance for the treatment and prognosis of chRCC patients to study influential prognostic factors, establish a high-precision prediction model and improve clinical treatment strategies for the particular group of the elderly, the central disease.

In recent years, with the broad application of the nomogram prediction model, traditional TNM staging has been gradually replaced. The prognosis of patients is also affected by many non-anatomic factors, such as age, gender, race, marital status, surgical methods, etc. (13). Nomogram is a data-based graphical computing tool that can estimate the risk of a disease based on staging systems such as the American Joint Commission on Cancer (AJCC) and other key risk factors related to prognosis (14). Today, many nomograms have been developed and applied to RCC, but mainly for patients with ccRCC and pRCC (15–18). However, there is still a lack of adequate and reliable nomograms to predict the prognosis of elderly patients with chRCC.

At present, artificial intelligence has been widely used in human health. Gadekallu et al. (19, 20) used deep learning to detect diabetic retinopathy. Kutia et al. (21) also discussed the great convenience that e-health systems bring to people. Kumar et al. (22) used neural networks to predict COVID-19. Iwendi et al. (23) use machine learning model for product recommendation. At present, there is no survival prediction model for chromophobe cell carcinoma in elderly patients. The existing prediction models are suitable for most of the population, but their accuracy is low.

Based on the above situation, we developed a competitive risk prediction model. We validated its accuracy in evaluating the cancer-specific survival rate (CSS) of elderly chRCC patients, providing a reference for clinical diagnosis and treatment.

METHODS

Data Source and Data Extraction

We collected clinicopathological data from patients in the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. The SEER database is the national cancer database of the United States, covering ∼28% of the US population and containing 18 cancer registries. Patient demographic information, clinicopathological information, and follow-up data can be obtained from the SEER database. Because the SEER database is a public database, we cannot identify patients' data, so our study does not require ethical approval and informed consent. Our methodology follows the rules of the SEER database.

We collected clinicopathological information for all chRCC patients, including age, sex, year of diagnosis, race, marital status, tumor size, laterality, histological tumor grade, TNM stage, surgical method, radiotherapy, chemotherapy, survival status, and cause of death. Inclusion criteria: (1) Pathological diagnosis of chromophobe renal cell carcinoma (ICD-O-3 code: 8317); (2) Age $\geq$ 65; (3) Unilateral renal tumor. Exclusion criteria; (1) TNM staging is unknown; (2) Unknown tumor size; (3) Unknown surgical method; (4) Survival time $<$1 month. The screening process for all patients is shown in **Figure 1**.

Patients were divided into years of diagnosis from 2004 to 2010 and 2011 to 2018. Patients were categorized as white, black, and other (American Indian /AK Native, Asian/Pacific Islander). Histological grades of tumors include grade I (highly differentiated), grade II (moderately differentiated), grade III (poorly differentiated), and grade IV (undifferentiated). Patients' surgeries were classified into three categories, including local tumor excision (codes 10–27), partial nephrectomy (codes 30), and radical nephrectomy (codes 40–80).

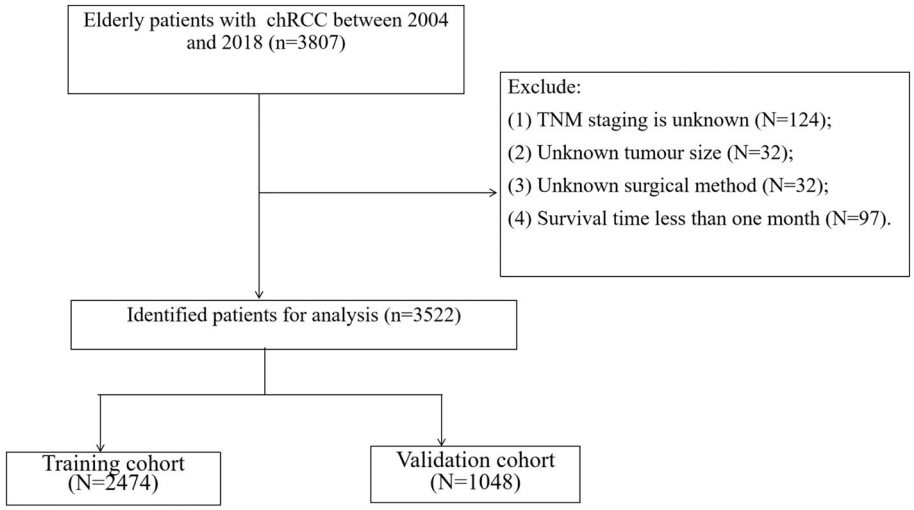


FIGURE 1 | Flow chart of patient screening.

Construction and Validation of the Competitive Risk Model

At present, the main research methods of competitive risk include cause-specific risk model and cumulative risk model. The cumulative risk model takes into account other competing endpoint events when calculating a related endpoint event, which is closer to reality (24). All patients were randomly assigned to a training cohort (70%) or a validation cohort (30%). In the training cohort, the primary outcome was cancer-specific mortality (CSM), and the competing risk event was other causes of mortality (OCM). In the case of competition risk, the cumulative risk model is used to estimate the cumulative occurrence rate of interest events. Based on the proportional sub-distribution hazard model proposed by Fine and Gray, the influencing factors of cancer-specific death in patients were analyzed. At the same time, we used the risk factors of CSM to construct a competitive risk model to predict CSS. Consistency index (C-index), the area under the receiver operating curve (AUC), calibration curve were used to validate the accuracy and discrimination of the model.

Clinical Utility

Decision curve analysis (DCA) was used to validate the clinical value of the model. We compared it with traditional TNM staging. All patients were divided into high-risk and low-risk groups based on competitive risk model scores. Log-rank test and Kaplan-Meier (K-M) curve were used to analyze survival differences between groups.

Statistical Analysis

Measurement data (age and tumor size) were described by mean and standard deviation, and a non-parametric *U*-test analyzed differences. Count data were characterized by frequency (%), and the Chi-square test performed difference analysis or non-parametric *U*-test. We analyzed risk factors for CSM and death from other causes with proportional subdistribution hazard (SH). Log-rank test and Kaplan-Meier curve were used to analyze the difference in survival between groups. All statistical methods were analyzed by R software 4.1.0 and SPSS 26.0. A P-value < 0.05 was considered statistically significant.

RESULTS

Clinical Features

A total of 3,522 elderly patients with chRCC were included in the analysis. Patients were randomly assigned to either the training cohort ($N = 2{,}474$) or the validation cohort ($N = 1{,}048$). The average age of the patients was 73.2 ± 6.11 years, and 2,907 (82.5%) were white, 2,035 (57.8%) were male, and 2,185 (62.0%) were married. There were 1,749 (49.7%) patients with tumors at the left side, and histological grades included 192 (5.45%) of grade I, 1,097 (31.1%) of grade II, 617 (17.5%) of Grade III, and 110 (3.12%) of grade IV. There were 2,974 (84.4%) patients in T1-T2, 3,470 (98.5%) patients in N0, and 3,452 (98.0%) patients in M0. The mean tumor size was 49.2 ± 35.0 mm. Local tumor excision was performed in 214 (6.08%) patients, partial nephrectomy was performed in 1,264 (35.9%) patients, and radical nephrectomy was performed in 1,903 (54.0%) patients. Fifty-five (1.56%) patients received chemotherapy and 23 (0.65%) received radiotherapy. The clinicopathological information of all patients is shown in **Table 1**. There was no significant difference between the training cohort and the validation cohort.

Construction of the Competing Risk Model

The training set analyzed the risk factors of CSM and OCM using SH and cumulative survival rate. The results showed that the risk factors of CSM in patients included age and race, T stage, N stage, M stage, tumor size, operation. Risk factors for OCM were age, sex, marital status, year of diagnosis, and tumor size (**Table 2**). We used patient risk factors for a cancer-specific

TABLE 1 | Clinicopathological characteristics of elderly patients with chRCC.

	ALL $N = 3,522$	Training cohort $N = 2,474$	Validation cohort $N = 1,048$	*p*
Age	73.2 (6.11)	73.2 (6.13)	73.2 (6.07)	0.733
Race				0.879
White	2,907 (82.5%)	2,041 (82.5%)	866 (82.6%)	
Black	458 (13.0%)	320 (12.9%)	138 (13.2%)	
Other	157 (4.46%)	113 (4.57%)	44 (4.20%)	
Sex				0.333
Male	2,035 (57.8%)	1,416 (57.2%)	619 (59.1%)	
Female	1,487 (42.2%)	1,058 (42.8%)	429 (40.9%)	
Marital				0.899
No	1,337 (38.0%)	937 (37.9%)	400 (38.2%)	
Married	2,185 (62.0%)	1,537 (62.1%)	648 (61.8%)	
Year of diagnosis				0.824
2004–2010	1,302 (37.0%)	918 (37.1%)	384 (36.6%)	
2010–2018	2,220 (63.0%)	1,556 (62.9%)	664 (63.4%)	
Laterality				0.341
Left	1,749 (49.7%)	1,242 (50.2%)	507 (48.4%)	
Right	1,773 (50.3%)	1,232 (49.8%)	541 (51.6%)	
Grade				0.801
I	192 (5.45%)	133 (5.38%)	59 (5.63%)	
II	1,097 (31.1%)	786 (31.8%)	311 (29.7%)	
III	617 (17.5%)	431 (17.4%)	186 (17.7%)	
IV	110 (3.12%)	78 (3.15%)	32 (3.05%)	
Unknown	1,506 (42.8%)	1,046 (42.3%)	460 (43.9%)	
T				0.076
T1-T2	2,974 (84.4%)	2,107 (85.2%)	867 (82.7%)	
T3-T4	548 (15.6%)	367 (14.8%)	181 (17.3%)	
N				0.993
N0	3,470 (98.5%)	2,438 (98.5%)	1,032 (98.5%)	
N1	52 (1.48%)	36 (1.46%)	16 (1.53%)	
M				1.000
M0	3,452 (98.0%)	2,425 (98.0%)	1,027 (98.0%)	
M1	70 (1.99%)	49 (1.98%)	21 (2.00%)	
Tumor size	49.2 (35.0)	49.0 (36.1)	49.6 (32.4)	0.615
Surgery				0.642
No	141 (4.00%)	101 (4.08%)	40 (3.82%)	
Local tumor excision	214 (6.08%)	143 (5.78%)	71 (6.77%)	
Partial nephrectomy	1,264 (35.9%)	897 (36.3%)	367 (35.0%)	
Radical nephrectomy	1,903 (54.0%)	1,333 (53.9%)	570 (54.4%)	
Chemotherapy				0.968
No/unknown	3,467 (98.4%)	2,436 (98.5%)	1,031 (98.4%)	
Yes	55 (1.56%)	38 (1.54%)	17 (1.62%)	
Radiation				0.539
No/Unknown	3,499 (99.3%)	2,456 (99.3%)	1,043 (99.5%)	
Yes	23 (0.65%)	18 (0.73%)	5 (0.48%)	

end to construct a competitive risk model that predicted patient CSS (**Figure 2**). We developed a web page utility to calculate a patient's probability of CSS. Visit https://chenghao.shinyapps.io/DynNomapp/ to enter the site and a patient's clinicopathological features to obtain a CSS rate.

Validation of the Competitive Risk Model

We use the C-index to validate the model's training and validation set accuracy. In the training set, the model predicted patients' 1-, 3-, and 5-year CSS with C-indices of 82.2, 80.8, and 78.2, respectively. The model predicted patient 1-, 3-, and 5-year CSS in the validation cohort with C-indices of 84.7, 83.4, and 76.9, respectively. The results show that the model has good discrimination. In addition, the model's calibration curve showed that the model's predicted value is almost consistent with the actual observed value, indicating that the model has good accuracy (**Figure 3**). The AUC of the training set and validation cohort also suggested that the model has good discrimination (**Figure 4**).

Clinical Application of the Competitive Risk Model

DCA showed that in both the training and validation set, the clinical utility of the model in predicting patients' 1-, 3-, and 5-year CSS was higher than that of traditional TNM staging (**Figure 5**). We classified patients into the high-risk group based on their competitive risk model score (total $>$ 54.1) and low-risk group (total score $\leq$ 54.1). K-M curve showed that the 1-, 3- and 5-year survival rates of patients in the high-risk group were 96.1, 92.3, and 88.2%, respectively, while the 1-, 3-, and 5-year survival rates of patients in the low-risk group were 99.1, 97.3, and 95.0%, respectively. In the training and validation sets, survival was significantly higher in the low-risk group than in the high-risk group (**Figure 6**). In addition, we analyzed surgical procedures in the high-risk and low-risk groups. In the low-risk group, all patients underwent surgery, and there were no significant differences in CSS between the types of surgery (**Figure 7A**). Most patients chose radical nephrectomy in the high-risk group, and patients with a partial nephrectomy and local tumor excision had significantly higher survival rates (**Figure 7B**).

DISCUSSION

chRCC is a new type of renal cell carcinoma (RCC) discovered by Thoenes et al. (25). The etiology of ChRCC is not clear yet, and some studies have claimed that it is closely related to chromosome variation. All chromosome loss is common in chRCCs, especially chromatids 1, 2, 6, 10, 13, and 17 (26). The inactivation of the TSC2 gene and the activation of MTOR is the most critical molecular changes (6). Under the pathological microscope, chRCC is characterized by large polygonal cells, pale cytoplasm, protruding cell membrane accompanied by raisin-like shrinkage (27). Ultrastructural analysis showed many abnormal mitochondria in the chRCC cytoplasm, with different sizes and shapes, accompanied by external swelling (28). In immunohistochemistry, CK7 is an important marker, expressed in more than 75% of chRCC, which can be preliminarily diagnosed by combining the diffuse positive expression of CD117 and KSP-cadherin (29).

TABLE 2 | Multivariate cox regression models predict cancer-specific mortality and other causes mortality in elderly patients with chRCC.

	CSM			OCM		
	HR	**95%CI**	***P***	**HR**	**95%CI**	***P***
Age	1.056	1.03–1.08	<0.001	1.076	1.06–1.09	<0.001
Race						
White						
Black	1.761	1.18–2.63	0.0055	1.238	0.95–1.62	0.12
Other	0.423	0.18–1	0.049	0.677	0.38–1.21	0.19
Sex						
Male						
Female	1.283	0.94–1.76	0.12	0.687	0.56–0.84	<0.001
Marital						
No						
Married	1.106	0.8–1.53	0.54	0.788	0.64–0.97	0.022
Year of diagnosis						
2004–2010						
2010–2018	0.772	0.56–1.06	0.11	0.786	0.63–0.97	0.028
Laterality						
Left						
Right	0.993	0.74–1.33	0.96	0.999	0.83–1.19	0.99
Grade						
I						
II	0.712	0.37–1.37	0.31	0.860	0.6–1.24	0.42
III	0.906	0.46–1.77	0.77	0.834	0.56–1.23	0.36
IV	1.850	0.85–4.02	0.12	0.586	0.3–1.14	0.12
Unknown	0.730	0.38–1.41	0.35	0.925	0.64–1.34	0.68
T						
T1-T2						
T3-T4	1.965	1.38–2.8	<0.001	0.975	0.75–1.28	0.85
N						
N0						
N1	4.552	2.38–8.71	<0.001	0.224	0.05–1.06	0.059
M						
M0						
M1	3.100	1.47–6.54	0.0029	0.522	0.12–2.22	0.38
Tumor size	1.004	1–1.01	<0.001	1.002	1–1	0.025
Surgery						
No						
Local tumor excision	0.346	0.15–0.82	0.017	0.920	0.47–1.79	0.81
Partial nephrectomy	0.197	0.11–0.37	<0.001	0.862	0.48–1.55	0.62
Radical nephrectomy	0.352	0.21–0.59	<0.001	0.855	0.48–1.52	0.59

However, it still needs to be differentiated from eosinophilic carcinoma. ChRCC has a clear boundary and no capsule, only occurs in the kidney, and the section is gray or light brown. Eosinophilic carcinoma can occur in any part of the body, and the cut surface is peach-red (30, 31). ChRCC lacks typical clinical manifestations. A characteristic triad of low back pain, hematuria and abdominal mass rarely occurs in the early stage of ChRCC and is considered a marker of disease progression (32).

At present, the nomogram of renal cell carcinoma has been widely established. For example, Wang et al. (15) established a nomogram to predict the medium- and long-term prognosis of patients with papillary cell carcinoma. Peng et al. (16) established a nomogram to predict the prognosis of patients with clear cell carcinoma and renal cell carcinoma. Yan et al. (17) established a nomogram to predict cancer-specific survival in patients with papillary renal cell carcinoma. Chen et al. established a nomogram to predict the prognosis of chromophobe RCC patients, but did not involve competitive risk. Although these nomgorams can be used to predict survival in patients with renal cell carcinoma, they are not accurate enough. The competitive risk of survival

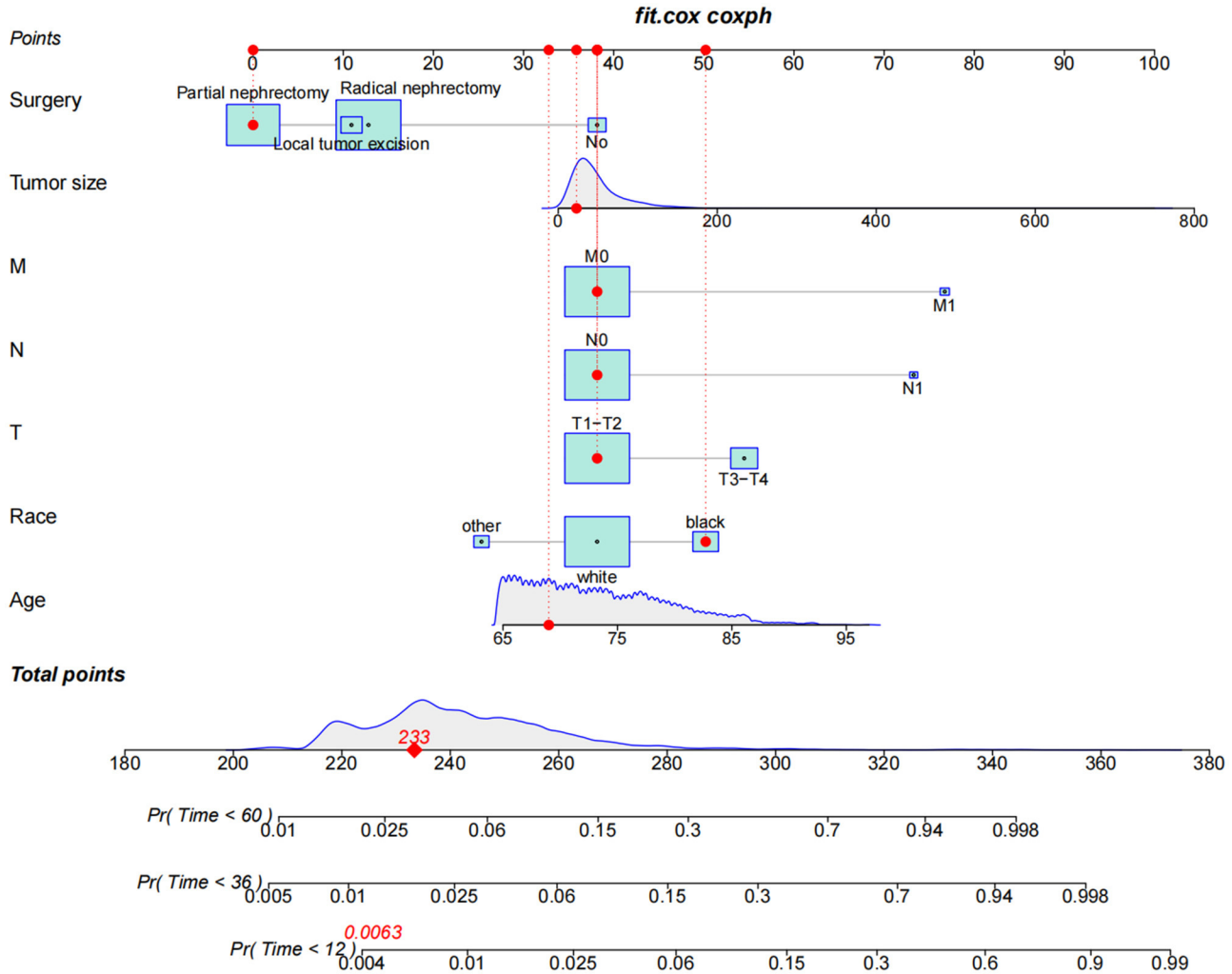


FIGURE 2 | The competitive risk model nomogram of CSS in elderly patients with chRCC at 1-, 3-, and 5-year.

in elderly patients with chRCC requires a more accurate prediction model.

The arrival of artificial intelligence makes big data medical care a reality. Maddikunta et al. (33) explore the potential application of artificial intelligence in industry 5.0. Healthcare 5.0 aims to use medical big data combined with artificial intelligence to assist human health. This study is to use the big data of cancer patients to establish a prediction model to predict the survival of patients.

Age is a critical factor in developing all cancers, and as we age, the risk of genetic mutations that could trigger cancer increases. Somatic mutations are generally considered the first step in the occurrence of cancer. They are associated with aging and highly reproducible DNA methylation changes, which help explain the higher prevalence of malignant tumors in the elderly (34). There is no consensus on defining the age of elderly patients, but more than 60% of initial cancer diagnoses and more than 70% of cancer deaths occur in patients over 65 years old (35). To improve the accuracy and representativeness of the prediction model, chRCC patients over 65 years old were included in this study. Escudier et al. (36) found that under the condition of the same stage and grade, there was a massive difference in the survival rate of patients and believed that clinical factors such as age were more critical than pathological and metastatic factors. Dias-Santos et al. (37) also showed that age is closely related to the survival rate of various cancers.

Surgery is a crucial measure for the treatment of chRCC, and tumor size based on clinical T-stage directly affects the choice of surgery and surgical method. Currently, the primary surgical procedures for chRCC include radical nephrectomy, partial nephrectomy, and local tumor excision represented by thermal ablation (TA), radiofrequency ablation (RFA), and cryoablation (CA). Current guidelines indicate that PN is still the standard recommended procedure for stage cT1a

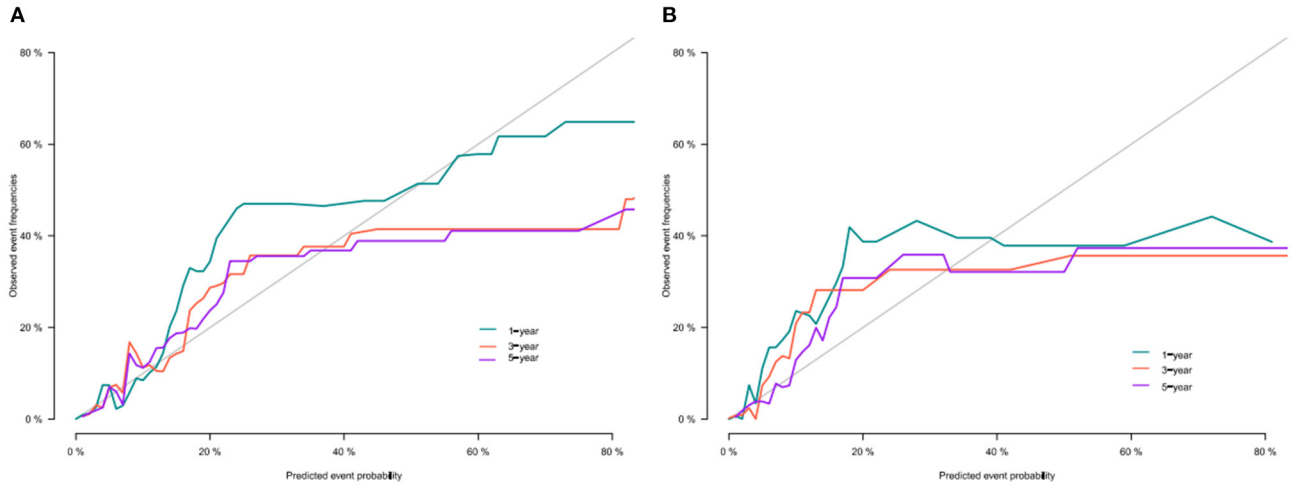


FIGURE 3 | Calibration curve of the nomogram in training cohort **(A)** and validation cohort **(B)**.

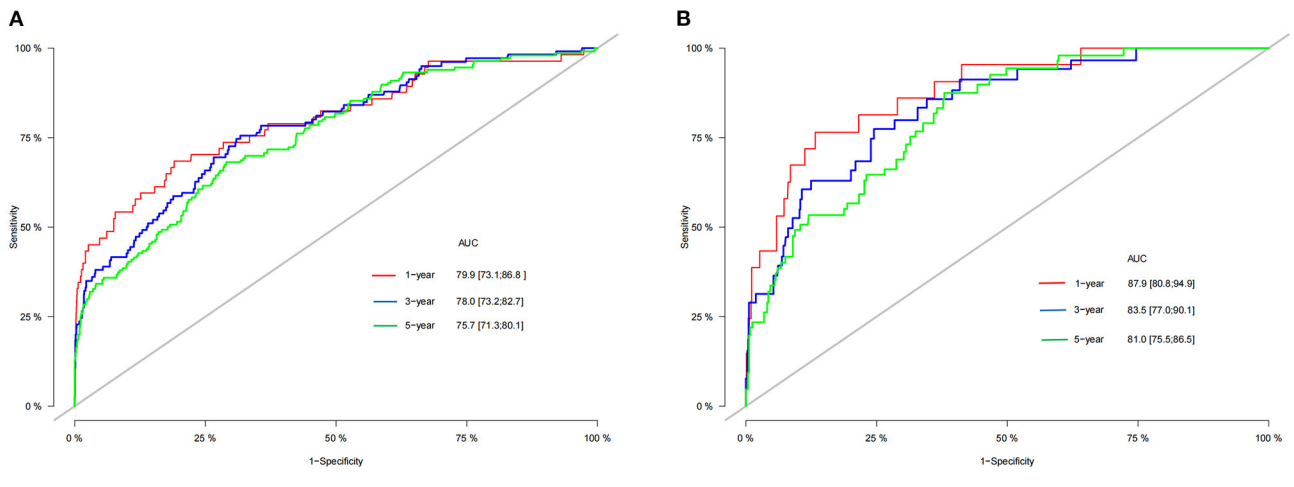


FIGURE 4 | AUC for predicting 1-, 3-, and 5-year CSS in training cohort **(A)** and validation cohort **(B)**.

tumors, while there is controversy for stage cT1b, and RN is recommended for stage cT2 (38). However, with the increasing reports of chronic kidney disease (CKD) induced by RN and PN surgery, more and more attention has been paid to the requirement of nephron preservation, and tumor resection represented by TA is recommended to be used in cT1a stage RCC (39). TA is theoretically feasible for tumors larger than 4 cm, but it requires multiple puncture operations at various sites, and the probability of bleeding and other complications is significantly increased (40). At the same time, based on the solid support of extensive sample prospective studies, active surveillance (AS) is recommended for renal tumor patients with cT1 stage (41), with metastasis rates of 0–6% and CSS of 0–18% (42). It is worth noting that Huang et al. (43) found that in elderly chRCC patients over 65 years old, the mortality rate of the RN group was significantly higher than the PN group, which was presumed to be due to the risk of surgery and anesthesia caused by cardiovascular and cerebrovascular diseases and chronic kidney disease. Thompson et al. (44) also found a significant positive correlation between RN and mortality; therefore, AS has particular advantages in elderly patients.

In addition, the mean tumor size of ChRCC was 6.0 cm, which was significantly more significant than the other subtypes of RCC. However, the degree of malignancy was not higher than that of other RCCS, indicating no correlation between tumor growth and malignancy among different RCC subtypes (45). However, for the chRCC subtype, there was a linear relationship between tumor size and recurrence (46), which was consistent with this study. This study showed that the larger the tumor,

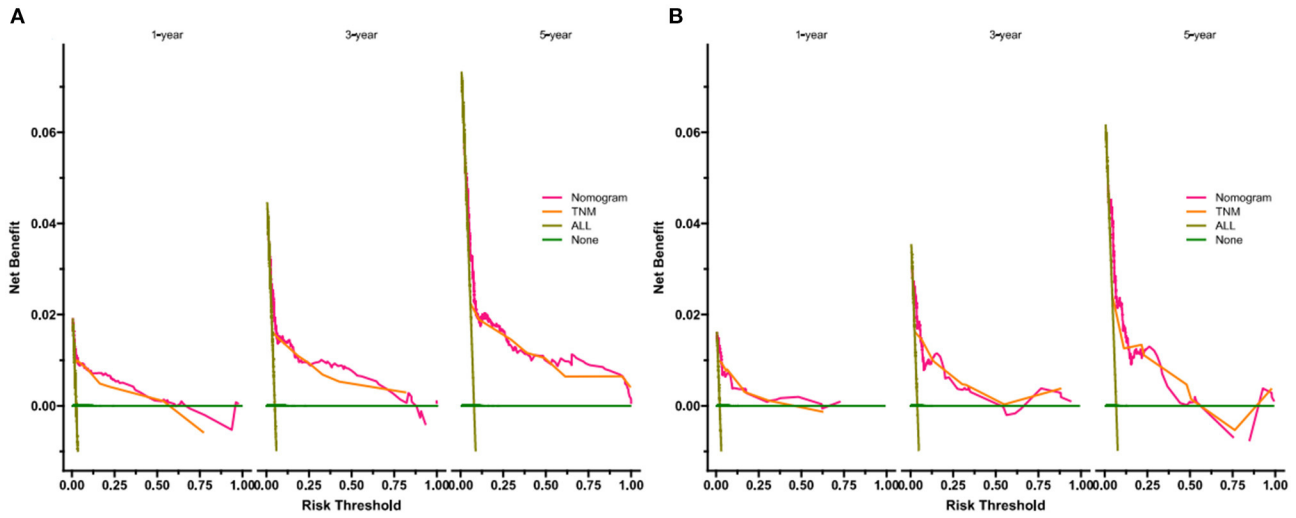


FIGURE 5 | DCA of the nomogram in training cohort **(A)** and validation cohort **(B)**. The Y-axis represents a net benefit and the X-axis represents threshold probability. The green line means no patients died and the dark green line means all patients died. When the threshold probability is between 20 and 100%, the net benefit of the model exceeds all deaths or none.

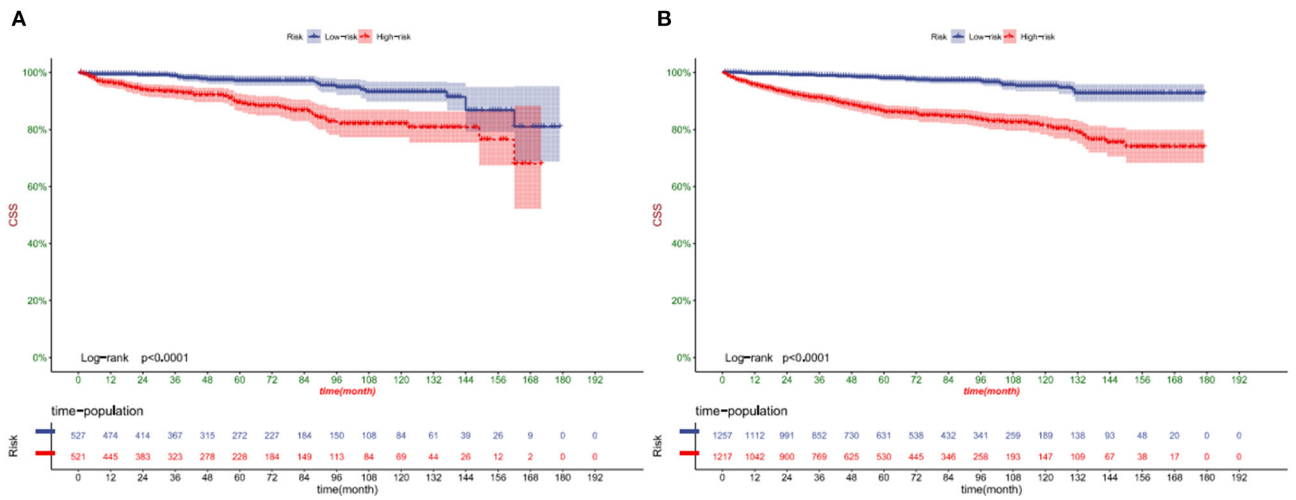


FIGURE 6 | Kaplan-Meier curves of patients in the low-risk and high-risk groups in training cohort **(A)** and validation cohort **(B)**.

the lower the patient's survival rate and could affect death from other causes.

This study is the first to construct a competitive risk model to predict cancer-specific survival in elderly patients with chromophobe cell carcinoma. This model has good accuracy and reliability through internal verification. However, there are still some shortcomings in our research. First of all, this study is a retrospective study, so it is challenging to avoid selection differences, which may cause a particular bias. Secondly, we did not include some key variables, such as BMI, smoking, drinking, hypertension, etc., which could improve the model's accuracy. Finally, our model has only undergone internal validation, and further external proof and future clinical application are necessary. Next, we will conduct the next prospective study to validate our prediction model.

CONCLUSION

We identified age, race, TNM stage, tumor size, and surgery as risk factors based on competitive risk in elderly patients with chRCC. We constructed a competitive risk model to predict CSS in elderly patients with chRCC. The model has good accuracy and reliability, which can help doctors and patients to make clinical decisions and follow-up strategies.

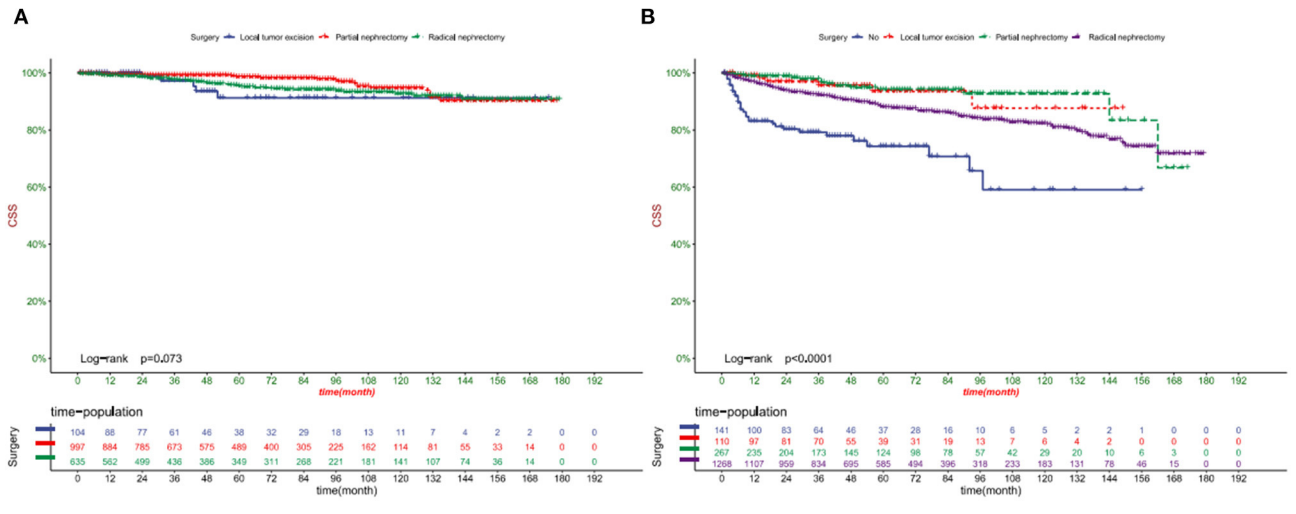


FIGURE 7 | Kaplan-Meier curves of patients with different surgical procedures in the low-risk group **(A)** and high-risk group **(B)**.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.cancer.gov/.

ETHICS STATEMENT

The data of this study is obtained from the SEER database. The patients' data is public and anonymous, so this study does not require ethical approval and informed consent.

AUTHOR CONTRIBUTIONS

JW and CZ designed the study. CZ, JW, JL, ML, LJ, and XT collected and analyzed the data. JW drafted the initial manuscript. CZ, TM, JL, ZZ, and DH revised the article critically. CZ, JL, DH, ML, and XT reviewed and edited the article. All authors approved the final manuscript.

FUNDING

This work was supported by Special Key Project of Chongqing Technology Innovation and Application Development (No. Cstc2019jscx-tjsbX0003).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Humphrey PA, Moch H, Cubilla AL, Ulbright TM, Reuter VE. The 2016 WHO classification of tumours of the urinary system and male genital organs-part B: prostate and bladder tumours. *Eur Urol.* (2016) 70:106–19. doi: 10.1016/j.eururo.2016.02.028
3. Badowska-Kozakiewicz AM, Budzik MP, Koczkodaj P, Przybylski J. Selected tumor markers in the routine diagnosis of chromophobe renal cell carcinoma. *Arch Med Sci.* (2016) 12:856–63. doi: 10.5114/aoms.2015.51188
4. Jia W, Hu C, Xu J, Ming J, Zhao Y, Cai M, et al. Dissolved organic matter derived from rape straw pretreated with selenium in soil improves the inhibition of Sclerotinia sclerotiorum growth. *J Hazard Mater.* (2019) 369:601–10. doi: 10.1016/j.jhazmat.2019.02.055
5. Maroto P, Anguera G, Roldan-Romero JM, Apellániz-Ruiz M, Algaba F, Boonman J, et al. Biallelic TSC2 mutations in a patient with chromophobe renal cell carcinoma showing extraordinary response to temsirolimus. *J Natl Compr Canc Netw.* (2018) 16:352–8. doi: 10.6004/jnccn.2017.7041
6. Moch H, Ohashi R. Chromophobe renal cell carcinoma: current and controversial issues. *Pathology.* (2021) 53:101–8. doi: 10.1016/j.pathol.2020.09.015
7. Zhou J, Yang X, Zhou L, Zhang P, Wang C. Combined immunohistochemistry for the "Three 7" markers (CK7, CD117, and Claudin-7) is useful in the diagnosis of chromophobe renal cell carcinoma and for the exclusion of mimics: diagnostic experience from a single institution. *Dis Markers.* (2019) 2019:4708154. doi: 10.1155/2019/4708154
8. Zini L, Leroy X, Lemaitre L, Devos P, Aubert S, Biserte J, et al. Tumour necrosis in chromophobe renal cell carcinoma: clinical data to distinguish aggressive variants. *Eur J Surg Oncol.* (2008) 34:687–91. doi: 10.1016/j.ejso.2007.08.010
9. Cheville JC, Lohse CM, Zincke H, Weaver AL, Blute ML. Comparisons of outcome and prognostic features among histologic subtypes of renal cell carcinoma. *Am J Surg Pathol.* (2003) 27:612–24. doi: 10.1097/00000478-200305000-00005
10. Bellmunt J, Puente J, Garcia de Muro J, Lainez N, Rodríguez C, Duran I, et al. SEOM clinical guidelines for the treatment of renal cell carcinoma. *Clin Transl Oncol.* (2014) 16:1043–50. doi: 10.1007/s12094-014-1219-1
11. Przybycin CG, Cronin AM, Darvishian F, Gopalan A, Al-Ahmadie HA, Fine SW, et al. Chromophobe renal cell carcinoma: a clinicopathologic study of 203 tumors in 200 patients with primary resection at a single institution. *Am J Surg Pathol.* (2011) 35:962–70. doi: 10.1097/PAS.0b013e31821a455d
12. Paner GP, Amin MB, Alvarado-Cabrero I, Young AN, Stricker HJ, Moch H, et al. novel tumor grading scheme for chromophobe renal cell carcinoma: prognostic utility and comparison with Fuhrman nuclear grade. *Am J Surg Pathol.* (2010) 34:1233–40. doi: 10.1097/PAS.0b013e3181e96f2a

13. Serdy KM, Leone JP, Dabbs DJ, Bhargava R. Male breast cancer. *Am J Clin Pathol.* (2017) 147:110–9. doi: 10.1093/ajcp/aqw207
14. Liu X, Wu Z, Lin E, Li W, Chen Y, Sun X, et al. Systemic prognostic score and nomogram based on inflammatory, nutritional and tumor markers predict cancer-specific survival in stage II-III gastric cancer patients with adjuvant chemotherapy. *Clin Nutr.* (2019) 38:1853–60. doi: 10.1016/j.clnu.2018.07.015
15. Wang H, Chen X, Zhao J, Kang M, Dong R, Wang K, et al. Predictive nomogram for midterm to long-term prognosis in patients with papillary renal cell carcinoma based on data from the surveillance, epidemiology, and end results (SEER) program. *Med Sci Monit.* (2020) 26:e921859. doi: 10.12659/MSM.921859
16. Peng Q, Zhou Y, Jin L, Cao C, Gao C, Zhou J, et al. Development and validation of an integrative methylation signature and nomogram for predicting survival in clear cell renal cell carcinoma. *Transl Androl Urol.* (2020) 9:1082–98. doi: 10.21037/tau-19-853
17. Yan H, Wei X, Wu A, Sha Y, Li X, Qi F. Nomograms for predicting overall and cancer-specific survival in patients with papillary renal cell carcinoma: a population-based study using SEER database. *Transl Androl Urol.* (2020) 9:1146–58. doi: 10.21037/tau-19-807
18. Chen C, Geng X, Liang R, Zhang D, Sun M, Zhang G, et al. Nomograms-based prediction of overall and cancer-specific survivals for patients with chromophobe renal cell carcinoma. *Exp Biol Med (Maywood).* (2021) 246:729–39. doi: 10.1177/1535370220977107
19. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics.* (2020) 9:274. doi: 10.3390/electronics9020274
20. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intellig Humanized Comput.* (2020) 1–14. doi: 10.1007/s12652-020-01963-7
21. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access.* (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
22. Kumar RL, Khan F, Din S, Band SS, Mosavi A, Ibeke E. Recurrent neural network and reinforcement learning model for COVID-19 prediction. *Front Public Health.* (2021) 9:744100. doi: 10.3389/fpubh.2021.744100
23. Iwendi C, Ibeke E, Eggoni H, Velagala S, Srivastava G. Pointer-based item-to-item collaborative filtering recommendation system using a machine learning model. *Int J Inform Technol Decision Making.* (2021) 1–22. doi: 10.1142/S0219622021500619
24. Glynn RJ, Rosner B, Christen WG. Evaluation of risk factors for cataract types in a competing risks framework. *Ophthalmic Epidemiol.* (2009) 16:98–106. doi: 10.1080/09286580902737532
25. Thoenes W, Störkel S, Rumpelt HJ. Human chromophobe cell renal carcinoma. *Virchows Arch B Cell Pathol Incl Mol Pathol.* (1985) 48:207–17. doi: 10.1007/BF02890129
26. Tan MH, Wong CF, Tan HL, Yang XJ, Ditlev J, Matsuda D, et al. Genomic expression and single-nucleotide polymorphism profiling discriminates chromophobe renal cell carcinoma and oncocytoma. *BMC Cancer.* (2010) 10:196. doi: 10.1186/1471-2407-10-196
27. Stec R, Grala B, Maczewski M, Bodnar L, Szczylik C. Chromophobe renal cell cancer–review of the literature and potential methods of treating metastatic disease. *J Exp Clin Cancer Res.* (2009) 28:134. doi: 10.1186/1756-9966-28-134
28. Rathmell KW, Chen F, Creighton CJ. Genomics of chromophobe renal cell carcinoma: implications from a rare tumor for pan-cancer studies. *Oncoscience.* (2015) 2:81–90. doi: 10.18632/oncoscience.130
29. Liu L, Qian J, Singh H, Meiers I, Zhou X, Bostwick DG. Immunohistochemical analysis of chromophobe renal cell carcinoma, renal oncocytoma, and clear cell carcinoma: an optimal and practical panel for differential diagnosis. *Arch Pathol Lab Med.* (2007) 131:1290–7. doi: 10.5858/2007-131-1290-IAOCRC
30. Michalova K, Tretiakova M, Pivovarcikova K, Alaghehbandan R, Perez Montiel D, Ulamec M, et al. Expanding the morphologic spectrum of chromophobe renal cell carcinoma: a study of 8 cases with papillary architecture. *Ann Diagn Pathol.* (2020) 44:151448. doi: 10.1016/j.anndiagpath.2019.151448
31. Ohashi R, Martignoni G, Hartmann A, Caliò A, Segala D, Stöhr C, et al. Multi-institutional re-evaluation of prognostic factors in chromophobe renal cell carcinoma: proposal of a novel two-tiered grading scheme. *Virchows Arch.* (2020) 476:409–18. doi: 10.1007/s00428-019-02710-w
32. Zbar B, Alvord WG, Glenn G, Turner M, Pavlovich CP, Schmidt L, et al. Risk of renal and colonic neoplasms and spontaneous pneumothorax in the Birt-Hogg-Dubé syndrome. *Cancer Epidemiol Biomarkers Prev.* (2002) 11:393–400.
33. Maddikunta PKR, Pham QV, Prabadevi B, Deepa N, Dev K, Gadekallu TR, et al. Industry 5.0: a survey on enabling technologies and potential applications. *J Indust Inform Integ.* (2021) 100257. doi: 10.1016/j.jii.2021.100257
34. Lin Q, Wagner W. Epigenetic aging signatures are coherently modified in cancer. *PLoS Genet.* (2015) 11:e1005334. doi: 10.1371/journal.pgen.1005334
35. González León T, Morera Pérez M. Renal cancer in the elderly. *Curr Urol Rep.* (2016) 17:6. doi: 10.1007/s11934-015-0562-2
36. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, et al. Electronic address: clinicalguidelines@esmoorg Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2019) 30:706–20. doi: 10.1093/annonc/mdz056
37. Dias-Santos D, Ferrone CR, Zheng H, Lillemoe KD, Fernández-Del Castillo C. The Charlson age comorbidity index predicts early mortality after surgery for pancreatic cancer. *Surgery.* (2015) 157:881–7. doi: 10.1016/j.surg.2014.12.006
38. Campbell S, Uzzo RG, Allaf ME, Bass EB, Cadeddu JA, Chang A, et al. Renal mass and localized renal cancer: AUA guideline. *J Urol.* (2017) 198:520–9. doi: 10.1016/j.juro.2017.04.100
39. Gunn AJ, Parikh NS, Bhatia S. Society of interventional radiology quality improvement standards on percutaneous ablation in renal cell carcinoma. *J Vasc Interv Radiol.* (2020) 31:195–201.e3. doi: 10.1016/j.jvir.2019.11.004
40. Welch BT, Shah PH, Thompson RH, Atwell TD. The current status of thermal ablation in the management of T1b renal masses. *Int J Hyperthermia.* (2019) 36:31–6. doi: 10.1080/02656736.2019.1605097
41. Uzosike AC, Patel HD, Alam R, Schwen ZR, Gupta M, Gorin MA, et al. Growth kinetics of small renal masses on active surveillance: variability and results from the DISSRM registry. *J Urol.* (2018) 199:641–8. doi: 10.1016/j.juro.2017.09.087
42. Finelli A, Cheung DC, Al-Matar A, Evans AJ, Morash CG, Pautler SE, et al. Small renal mass surveillance: histology-specific growth rates in a biopsy-characterized cohort. *Eur Urol.* (2020) 78:460–7. doi: 10.1016/j.eururo.2020.06.053
43. Huang WC, Elkin EB, Levey AS, Jang TL, Russo P. Partial nephrectomy versus radical nephrectomy in patients with small renal tumors–is there a difference in mortality and cardiovascular outcomes? *J Urol.* (2009) 181:55–61; discussion 61–2. doi: 10.1016/j.juro.2008.09.017
44. Thompson RH, Boorjian SA, Lohse CM, Leibovich BC, Kwon ED, Cheville JC, et al. Radical nephrectomy for pT1a renal masses may be associated with decreased overall survival compared with partial nephrectomy. *J Urol.* (2008) 179:468–71; discussion 472–3. doi: 10.1016/j.juro.2007.09.077
45. Kawaguchi S, Fernandes KA, Finelli A, Robinette M, Fleshner N, Jewett MA. Most renal oncocytomas appear to grow: observations of tumor kinetics with active surveillance. *J Urol.* (2011) 186:1218–22. doi: 10.1016/j.juro.2011.05.080
46. Casuscelli J, Becerra MF, Seier K, Manley BJ, Benfante N, Redzematovic A, et al. Chromophobe renal cell carcinoma: results from a large single-institution series. *Clin Genitourin Cancer.* (2019) 17:373–9.e4. doi: 10.1016/j.clgc.2019.06.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Zhanghuang, Tan, Mi, Liu, Jin, Li, Zhang and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 28 February 2022
doi: 10.3389/fpubh.2022.848716

Development and Validation of a Nomogram to Predict Cancer-Specific Survival for Middle-Aged Patients With Early-Stage Hepatocellular Carcinoma

Chong Wen[1,2†], *Jie Tang*[3†] *and Hao Luo*[1*]

[1] General Surgery Center, The General Hospital of Western Theater, Chengdu, China, [2] College of Medicine, Southwest Jiaotong University, Chengdu, China, [3] Department of Biostatistics and Epidemiology, School of Public Health, Shenyang Medical College, Shenyang, China

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Senthilkumar Mohan,
VIT University, India
Ebuka Ibeke,
Robert Gordon University,
United Kingdom

***Correspondence:**
Hao Luo
luohao@swjtu.edu.cn

[†] These authors share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 05 January 2022
Accepted: 07 February 2022
Published: 28 February 2022

Citation:
Wen C, Tang J and Luo H (2022) Development and Validation of a Nomogram to Predict Cancer-Specific Survival for Middle-Aged Patients With Early-Stage Hepatocellular Carcinoma. Front. Public Health 10:848716. doi: 10.3389/fpubh.2022.848716

Background: Hepatocellular carcinoma is a common cause of death in middle-aged patients. We aimed to construct a new nomogram to predict cancer-specific survival (CSS) in middle-aged patients with hepatocellular carcinoma at an early stage.

Method: We collected clinicopathological information on early middle-aged patients with hepatocellular carcinoma from the SEER database. Univariate and multivariate Cox regression models were used to screen the independent risk factors for prognosis. These risk factors were used to construct predictions of CSS in patients with hepatocellular carcinoma. Consistency index (C- index), calibration curve, area under the receiver operating curve (AUC) were used. A decision analysis curve (DCA) was used to evaluate the clinical utility of the predictive model.

Results: A total of 6,286 patients with hepatocellular carcinoma in early middle age were enrolled. Univariate and multivariate Cox regression analysis showed that sex, marriage, race, histological tumor grade, T stage, surgery, chemotherapy, AFP, and tumor size were independent risk factors for prognosis. All independent risk factors were included in the nomogram to predict CSS at 1-, 3-, and 5-years in early middle age patients with hepatocellular carcinoma. In the training cohort and validation cohort, the C-index of the prediction model was 0.728 (95%CI: 0.716–0.740) and 0.733 (95%CI: 0.715–0.751), respectively. The calibration curve showed that the predicted value of the prediction model is highly consistent with the observed value. AUC also suggested that the model has good discrimination. DCA suggested that the nomogram had better predictive power than T staging.

Conclusion: We constructed a new nomogram to predict CSS in middle-aged patients with early-stage hepatocellular carcinoma. This prediction model has good accuracy and reliability, which can help patients and doctors to judge prognosis and make clinical decisions.

Keywords: nomogram, cancer-specific survival, middle-aged patients, hepatocellular carcinoma, SEER

INTRODUCTION

Primary liver cancer is the sixth most commonly diagnosed cancer and the third leading cause of cancer-related death worldwide, with ∼906,000 new cases and deaths in 2020, and hepatocellular carcinoma (HCC) accounts for the vast majority (75–85%) of primary liver cancers (1). Although some risk factors such as hepatitis B virus (HBV) and aflatoxin B1 (AFB1) are declining in many regions due to public health measures, the incidence of liver cancer is still rising, possibly because of the increasing prevalence of obesity and diabetes (2). Due to a variety of diagnostic methods, including ultrasonography, serum alpha-fetoprotein (AFP), computed tomography (CT), and magnetic resonance imaging (MRI) being widely used, a more significant proportion of early HCC patients are diagnosed (3). Besides, a majority of HCC patients are aged between 40 and 60. Hence, the survival prognosis of this group of patients with early HCC remains a significant concern.

Many risk factors are affecting the survival prognosis in patients with HCC, including age (4), race (5), histological grade (6), chemotherapy (7), surgery (8), tumor size (9), and so on. To the best of our knowledge, numerous nomogram prediction models based upon the above factors for HCC have been constructed (10–12), as well as prognostic nomogram for early-stage HCC (13). However, there have been no clinical studies and statistical tools that evaluated the predictive factors for early-stage HCC patients aged between 40 and 60. Considering the earlier the stage and the younger the patient, the better the prognosis and the more meaningful the prediction, we aim to extract data from the Surveillance, Epidemiology, and End Results (SEER) database to develop and validate a practical nomogram for predicting the cancer-specific survival (CSS) of patients with early HCC aged between 40 and 60 by integrating some significant variables so that clinicians can make better decisions.

At present, artificial intelligence has been widely used in human health care. Shubham et al. (14) used deep learning methods to identify glomerulus in human kidney tissue images. Movassagh et al. (15) used a new method to train neural networks. Mohan et al. (16) used machine learning to predict heart disease. Iwendi et al. (17, 18) used machine learning to classify COVID-19 patients and make health predictions. Kumar et al. (19) used neural networks to predict COVID-19. Ngabo et al. (20) use machine learning architectures to tackle urban pandemics. In this study, we used conventional nomogram to construct a new predictive model to predict the survival of middle-aged HCC patients. This prediction model can help HCC patients and doctors to develop clinical and follow-up strategies.

PATIENTS AND METHODS

Data Source and Data Extraction

We downloaded clinicopathological information of all patients with HCC from 2010 to 2018 from the SEER database. We collected social and demographic information (age, sex, race, year of diagnosis, and marriage), tumor information (tumor stage, tumor profile, tumor size, and histological grade), and treatment information (radiation, chemotherapy, and surgery). Because SEER is a public database, patient information is not identifiable, so our study does not require ethical approval and patients' informed consent. Our analysis follows the data usage rules of the SEER database.

Inclusion criteria: (1) Pathological diagnosis of HCC; (2) Aged 40–60; (3) The years of diagnosis were 2010–2018. Exclusion criteria: (1) Surgical method unknown; (2) Unknown tumor size; (3) The cause of death is unknown; (4) Survival time $<$1 month. The patient inclusion and exclusion procedures are shown in **Figure 1**.

Patients were categorized as white, black, and others (American Indian/AK Native, Asian/Pacific Islander). The histological tumor grading of the patients was divided into grades I to IV, which were highly differentiated, moderately differentiated, poorly differentiated, and undifferentiated. The surgical methods of patients are divided into three categories: local tumor destruction, lobectomy, and transplant. AFP can be positive or negative.

Construction and Validation of the Nomogram

All patients were randomly assigned to a training cohort (70%) or a validation cohort (30%). Univariate Cox regression analysis was used to identify prognostic risk factors in middle-aged patients with early-stage HCC in the training cohort. Multivariate Cox regression analysis analyzed the independent risk factors for prognosis. All independent risk factors were included in a nomogram to predict patients' CSS. A series of validation methods were used to validate the accuracy and discrimination of the prediction model, including Consistency index (C-index), calibration curve, area under the receiver operating curve (AUC).

Clinical Utility

A decision analysis curve (DCA) is a method to evaluate the practical value of the model by calculating the net income under different thresholds. We used DCA to assess the clinical utility of the nomogram prediction model and compared it with conventional T staging. At the same time, we divided all patients into low-risk and high-risk groups based on each patient's nomogram score. Log-rank test and Kaplan-Meier (K-M) curve were used to compare survival differences among patients in different groups.

Statistical Analysis

Counting data were expressed as frequency (%), and differences between groups were analyzed by Chi-square or non-parametric U-tests. Measurement data (age and tumor size) were described using mean and standard deviation, and differences between groups were analyzed using non-parametric U-tests. Univariate and multivariate Cox regression models examined the survival and prognostic factors of the patients. Log-rank test and Kaplan-Meier (K-M) curve were used to compare survival differences between groups. R software 4.1.0 and SPSS 26.0 were used for statistical analysis. A P-value < 0.05 was considered statistically significant.

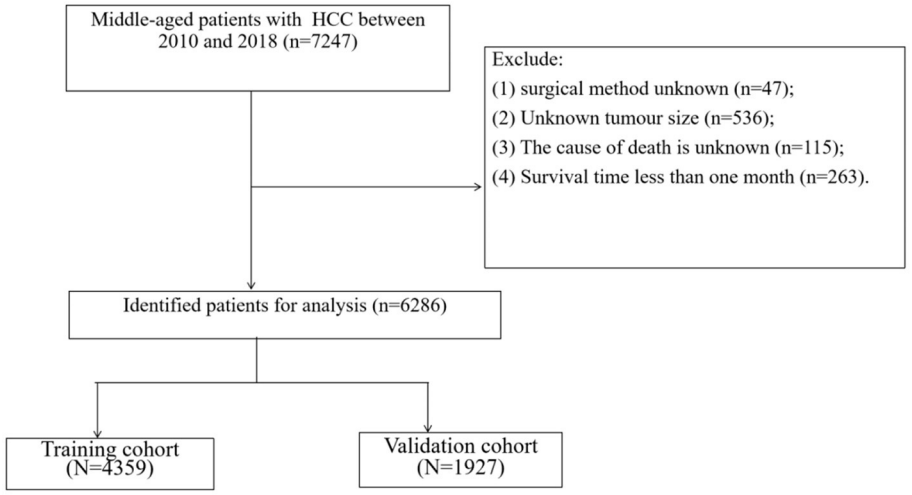


FIGURE 1 | Flow chart of patient screening.

RESULTS

Clinical Features

A total of 6,286 patients with early-stage HCC aged between 40 and 60 were included in the study. The mean age of all patients was 54.4 ± 4.1 years, 4,298 (68.4%) patients were white, 5,005 (79.6%) patients were male, 2,799 (44.5%) patients were married, and 3,386 (53.9%) patients were AFP positive. The histological grades were 666 (10.6%) in grade I, 1,050 (16.7%) in grade II, 309 (4.92%) in grade III, and 16 (0.25%) in grade IV. There were 3,976 patients (63.3%) with stage T1 tumors, with a mean tumor size of 35.5 ± 32.2 mm. Partial tumor destruction was performed in 1,183 (18.8%) patients, lobectomy was performed in 702 (11.2%) patients, and liver transplantation was performed in 769 (12.2%) patients. Two thousand nine hundred and fourteen (46.4%) patients received chemotherapy, and 507 (8.07%) received radiotherapy. The clinicopathological information of all patients is shown in **Table 1**. There was no significant difference between the training cohort and the validation cohort.

Univariate and Multivariate Cox Regression Analysis

In the training cohort, univariate Cox regression analysis suggested that sex, marriage, year of diagnosis, race, histological tumor grade, T stage, surgery, chemotherapy, AFP, and tumor size were correlated with prognosis. All related risk factors were included in multivariate Cox regression analysis. The results showed that sex, marriage, race, histological tumor grade, T stage, surgery, chemotherapy, AFP, and tumor size were independent risk factors for patient prognosis. The univariate and multivariate results are shown in **Table 2**.

Construction and Validation of the Nomogram

All independent risk factors were included in the nomogram to predict 1-, 3-, and 5-year CSS in early middle-aged HCC patients (**Figure 2**). Tumor size and operation were the most important factors influencing patient survival, followed by histological tumor grade, AFP, and T stage. In addition, race, marriage, and sex also influenced the prognosis of patients. In the training and validation cohort, the C-index of the prediction model was 0.728 (95%CI, 0.716–0.740) and 0.733 (95%CI, 0.715–0.751), respectively, indicating that the model had good discrimination. Calibration curves of the training cohort and validation cohort showed that the predicted value of the prediction model is highly consistent with the actual observed value (**Figure 3**), indicating that the model has good accuracy. In addition, the AUC of the training cohort and validation cohort also suggested that the model had good discrimination (**Figure 4**).

Clinical Application of the Nomogram

DCA suggested that the nomogram had better predictive power than the T-staging (**Figure 5**). All patients were assigned to the high-risk group based on their nomogram scores (total >133.2) and low-risk group (total ≤133.2). In the training and validation cohort, patients in the low-risk group had significantly higher survival rates than those in the high-risk group (**Figure 6**). In the high-risk group, the 1-, 3-, and 5-year survival rates were 78.2, 53.5, and 43.3%, respectively. In the low-risk group, the 1-, 3-, and 5-year survival rates were 84.7, 67.9, and 59.8%, respectively. In the low-risk group, survival was highest among patients who received liver transplantation and lobectomy (**Figure 7A**). Survival of patients who received liver transplantation was highest in the high-risk group (**Figure 7B**).

TABLE 1 | Clinicopathological characteristics of patients with HCC.

	All *N* = 6,286	Training cohort *N* = 4,359	Validation cohort 2*N* = 1,927	*p*
Age	54.4 (4.10)	54.4 (4.11)	54.4 (4.10)	0.810
Race				0.058
White	4,298 (68.4%)	2,940 (67.4%)	1,358 (70.5%)	
Black	1,004 (16.0%)	715 (16.4%)	289 (15.0%)	
Other	984 (15.7%)	704 (16.2%)	280 (14.5%)	
Sex				0.447
Male	5,005 (79.6%)	3,459 (79.4%)	1,546 (80.2%)	
Female	1,281 (20.4%)	900 (20.6%)	381 (19.8%)	
Year of diagnosis				0.762
2010–2014	3,928 (62.5%)	2,718 (62.4%)	1,210 (62.8%)	
2015–2018	2,358 (37.5%)	1,641 (37.6%)	717 (37.2%)	
Marriage				0.845
No	3,487 (55.5%)	2,414 (55.4%)	1,073 (55.7%)	
Married	2,799 (44.5%)	1,945 (44.6%)	854 (44.3%)	
AFP				0.853
Negative	1,848 (29.4%)	1,279 (29.3%)	569 (29.5%)	
Positive	3,386 (53.9%)	2,357 (54.1%)	1,029 (53.4%)	
Unknown	1,052 (16.7%)	723 (16.6%)	329 (17.1%)	
Grade				0.818
I	666 (10.6%)	453 (10.4%)	213 (11.1%)	
II	1,050 (16.7%)	725 (16.6%)	325 (16.9%)	
III	309 (4.92%)	211 (4.84%)	98 (5.09%)	
IV	16 (0.25%)	10 (0.23%)	6 (0.31%)	
Unknown	4,245 (67.5%)	2,960 (67.9%)	1,285 (66.7%)	
T				0.162
T1	3,976 (63.3%)	2,732 (62.7%)	1,244 (64.6%)	
T2	2,310 (36.7%)	1,627 (37.3%)	683 (35.4%)	
Tumor size	35.5 (32.2)	35.4 (32.0)	35.6 (32.7)	0.830
Surgery				0.325
No	3,632 (57.8%)	2,549 (58.5%)	1,083 (56.2%)	
Local tumor destruction	1,183 (18.8%)	816 (18.7%)	367 (19.0%)	
Lobectomy	702 (11.2%)	473 (10.9%)	229 (11.9%)	
Transplant	769 (12.2%)	521 (12.0%)	248 (12.9%)	
Chemotherapy				0.965
No/unknown	3,372 (53.6%)	2,337 (53.6%)	1,035 (53.7%)	
Yes	2,914 (46.4%)	2,022 (46.4%)	892 (46.3%)	
Radiation				0.370
No/unknown	5,779 (91.9%)	3,998 (91.7%)	1,781 (92.4%)	
Yes	507 (8.07%)	361 (8.28%)	146 (7.58%)	

DISCUSSION

Predictive models for cancer patients have been widely established, including nomograms for predicting survival in HCC patients. Yan et al. (13) constructed a nomogram to predict cancer-specific survival in early hepatocellular patients. Kong et al. (21) constructed a nomogram to predict survival in young patients with hepatocellular carcinoma. Hu et al. (22) developed a nomogram to predict survival in bone metastatic hepatocellular carcinoma patients. Liu et al. (23) constructed a nomogram to predict cancer-specific survival in all patients with hepatocellular carcinoma. Similarly, Ni et al. (24) constructed nomograms to predict survival in all hepatocellular patients. Nevertheless, prognosis and risk of death in HCC patients vary greatly with age. Middle-aged people have a high incidence of hepatocellular carcinoma, and their prognostic survival factors are different from those of other age groups. For example, marriage and race can affect the survival of patients, so a special nomogram is needed to predict patient survival.

TABLE 2 | Univariate and multivariate analyses of CSS in training cohort.

	Univariate			Multivariate		
	HR	**95%CI**	***P***	**HR**	**95%CI**	***P***
Age	1.01	0.99–1.02	0.355			
Race						
White	Reference			Reference		
Black	1.16	1.03–1.31	0.017	1.035	0.913–1.173	0.589
Other	0.63	0.55–0.74	<0.001	0.669	0.575–0.777	<0.001
Sex						
Male	Reference			Reference		
Female	0.82	0.73–0.93	0.002	0.013	0.855–0.756	
Year of diagnosis						
2010–2014	Reference			Reference		
2015–2018	0.88	0.79–0.98	0.024	0.906	0.809–1.013	0.084
Marriage						
No	Reference			Reference		
Married	0.66	0.6–0.73	<0.001	0.786	0.712–0.869	<0.001
AFP						
Negative	Reference			Reference		
Positive	1.87	1.66–2.1	<0.001	1.661	1.475–1.872	<0.001
Unknown	1.25	1.07–1.47	0.006	1.225	1.044–1.438	0.013
Grade						
I	Reference			Reference		
II	1.14	0.92–1.42	0.231	1.348	1.079–1.684	0.009
III	2.29	1.77–2.96	<0.001	1.987	1.529–2.583	<0.001
IV	1.78	0.66–4.82	0.256	3.531	1.284–9.71	0.014
Unknown	2.08	1.74–2.5	<0.001	1.155	0.959–1.391	0.129
T						
T1	Reference			Reference		
T2	1.14	0.92–1.42	0.231	1.511	1.372–1.663	<0.001
Tumor size	1	1–1.001	<0.001	1.003	1.002–1.003	<0.001
Surgery						
No	Reference			Reference		
Local tumor destruction	0.48	0.42–0.55	<0.001	0.469	0.408–0.538	<0.001
Lobectomy	0.34	0.28–0.41	<0.001	0.279	0.225–0.347	<0.001
Transplant	0.08	0.06–0.11	<0.001	0.074	0.053–0.103	<0.001
Chemotherapy						
No/unknown	Reference			Reference		
Yes	1.36	1.16–1.61	<0.001	0.757	0.685–0.837	<0.001
Radiation						
No/unknown	Reference					
Yes	1.14	1.03–1.25	0.008			

Our study extracted information on early-stage HCC patients aged between 40 and 60 from the SEER database. We split the entire cohort into training and validation groups. Differences between the two groups regarding various essential variables such as marriage status, AFP level, grade, stage, and treatments were tested. All tests showed p-values > 0.05, indicating no significant difference between the training and validation groups. Univariate and multivariate Cox analysis found that numerous factors significantly affected CSS, including sex, race, marital status, AFP level, grade, T stage, tumor size, surgery, and chemotherapy. At the same time, other variables such as age and year of diagnosis were not identified as prognostic significance. That is partly because the population we selected is strictly defined, and there was so little difference between patients about age and year of diagnosis.

Social and demographic variables (age, race, and marital status) were identified as prognosticators of HCC. In our study, marital status is a relatively important factor for survival prediction, parallel to many previous studies (13, 25). This is partly because most single, separated, and

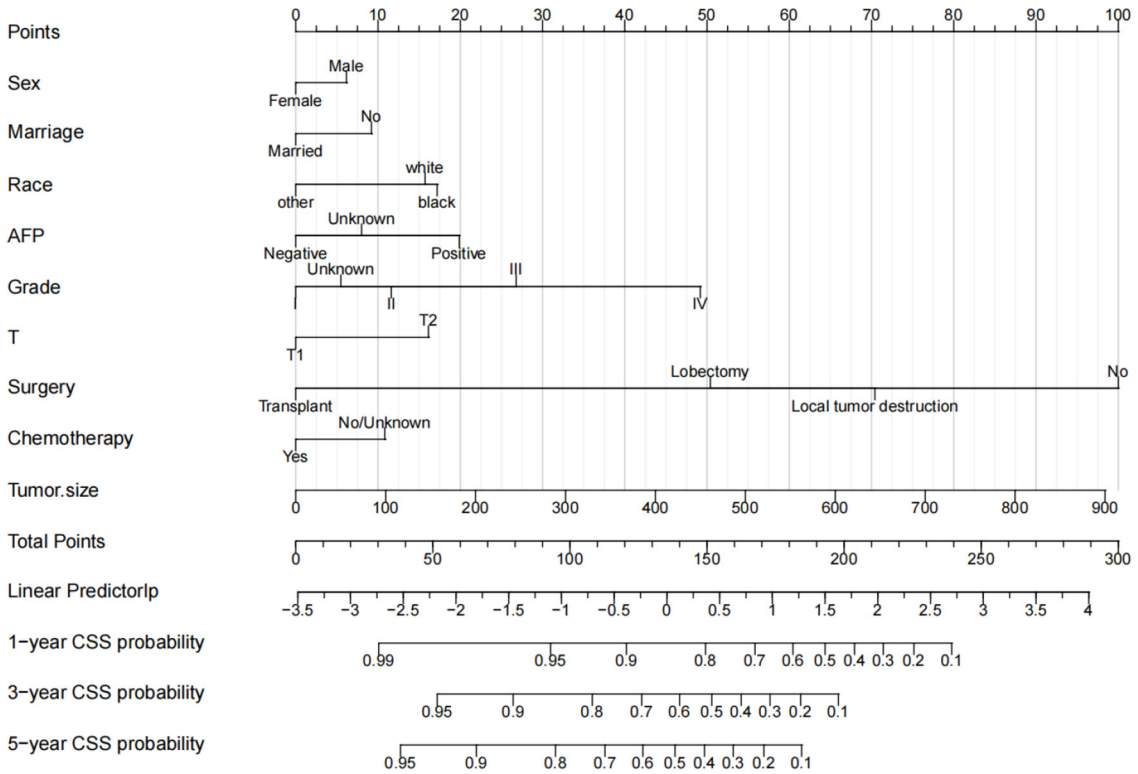


FIGURE 2 | The nomogram of CSS in middle-aged patients with HCC at 1-, 3-, and 5-year.

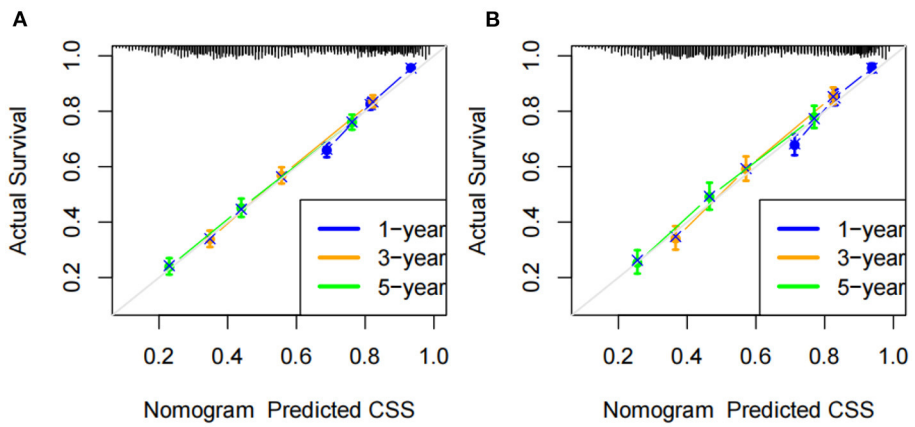


FIGURE 3 | Calibration curve of the nomogram. **(A)** Calibration curves of 1-, 3-, and 5-year CSS in training cohort; **(B)** calibration curves of 1-, 3-, and 5-year CSS in the validation cohort.

divorced cancer patients experience more stress and pain than married patients (26). Besides, married patients have better adherence to prescribed treatments than unmarried patients, leading to better cancer control (27). Furthermore, several biological, psychological, and social theories can also explain this phenomenon. A systematic review and meta-analysis indicated significant racial disparities in HCC prognosis in the United States, with Black patients having worse overall survival and Hispanic and Asian patients having better overall survival than White patients (28). Nikita Sandeep Wagle et al. found that Black patients receive curative treatment more minor than White because of social and economic situations (29), which may partly explain this phenomenon. Consistently, we found Black patients having worse CSS, closely followed by the White ($P = 0.589$), and other races having better CSS ($P < 0.001$; **Table 2**).

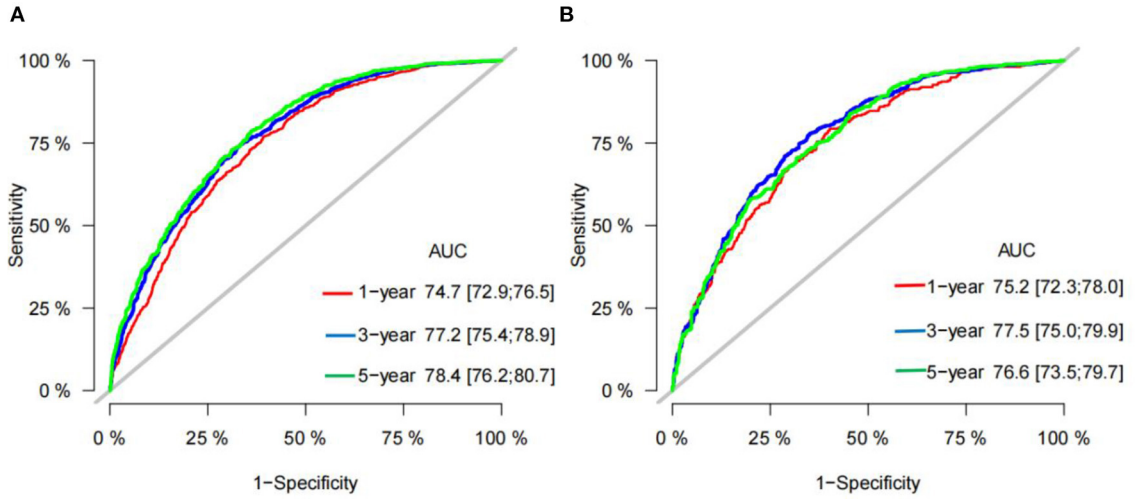


FIGURE 4 | AUC for predicting 1-, 3-, and 5-year CSS in training cohort **(A)** and validation cohort **(B)**.

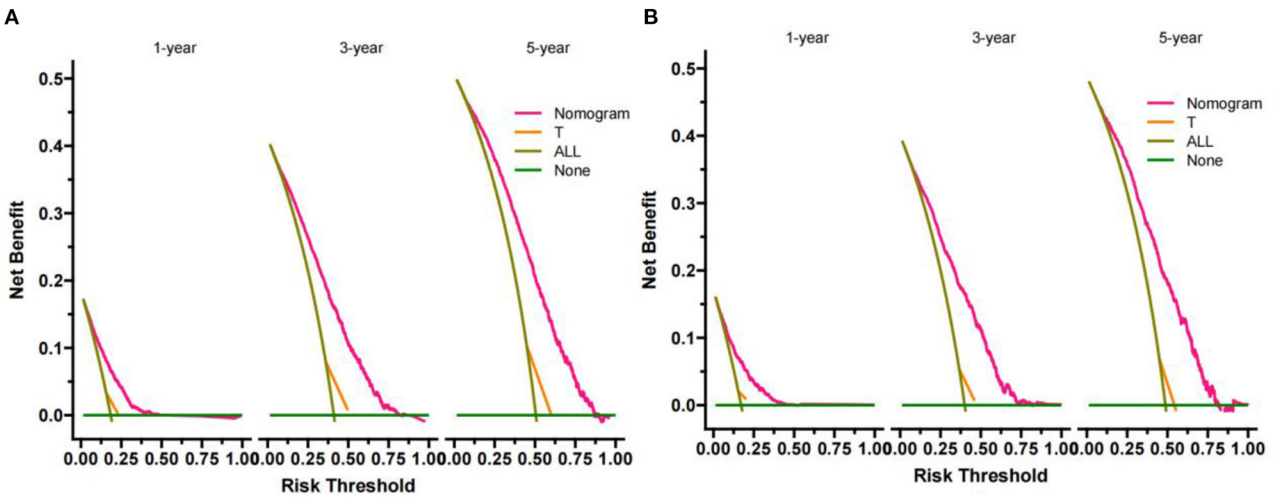


FIGURE 5 | DCA of the nomogram in training cohort **(A)** and validation cohort **(B)**. The Y-axis represents a net benefit, and the X-axis represents threshold probability. The green line means no patients died, and the dark green line means all patients died. When the threshold probability is between 0 and 75%, the net benefit of the model exceeds all deaths or none.

As for the tumor features (pathological grade, T stage, and tumor size) variables included in the nomogram, the tumor size has the most significant effect for CSS. A retrospective study suggests that tumor size is a vital independent prognostic factor in T2 solitary HCC after curative liver resection (30). In addition, the Milan criteria of liver transplantation strictly limited tumor size, which means that tumors with smaller diameters can be treated in more ways and get better treatment (31). Our results consistently reflected the relationship between tumor size and CSS and quantified its impact on CSS of a specified population compared with previous studies.

There are several kinds of treatment methods for patients with early-stage HCC, including liver resection (LR), ablation, and orthotopic liver transplantation (OLT) (32). Liver transplantation is considered an ideal treatment for selected patients with HCC, as it removes both the tumor and potential cirrhosis, with a 5-year survival rate exceeding 70% (33). Our results once again confirmed its superiority in the treatment of early-stage HCC, and its score is the lowest in the nomogram, which indicated the best outcome. However, the shortage of donor organs has severely limited the utilization of liver transplantation (34). In our results, lobectomy is inferior to liver transplantation but superior to local tumor destruction such as ablation. A systematic review of randomized controlled trials with Meta-analysis and sequential trial analysis found no statistical difference between radiofrequency ablation (RFA) and hepatic resection (HR) in

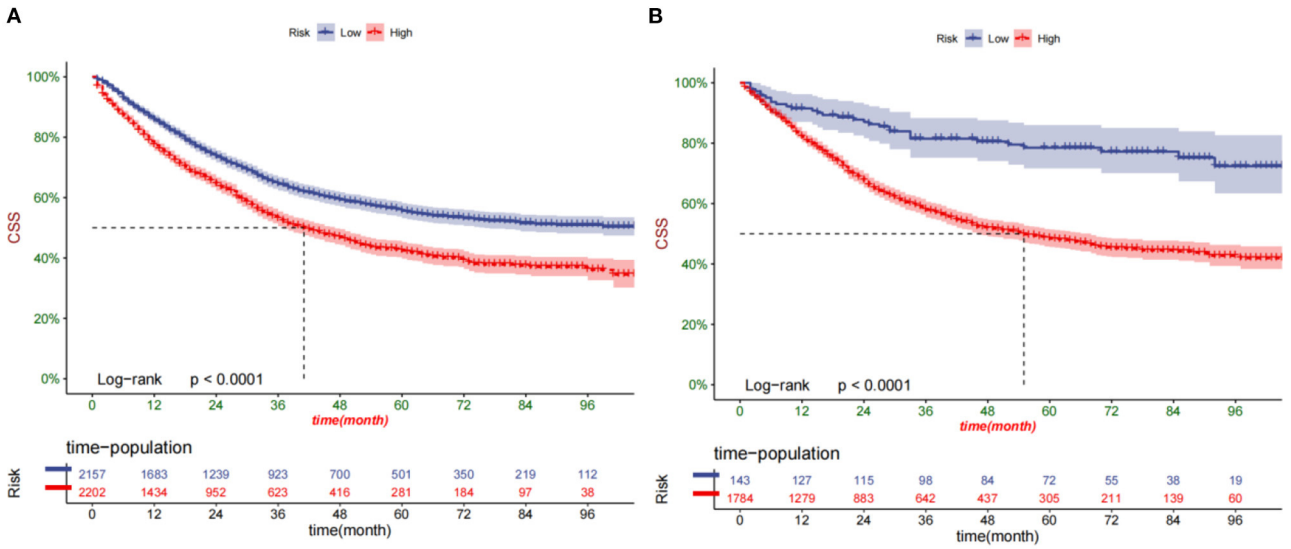


FIGURE 6 | Kaplan-Meier curves of patients in the low-risk and high-risk groups in training cohort **(A)** and validation cohort **(B)**.

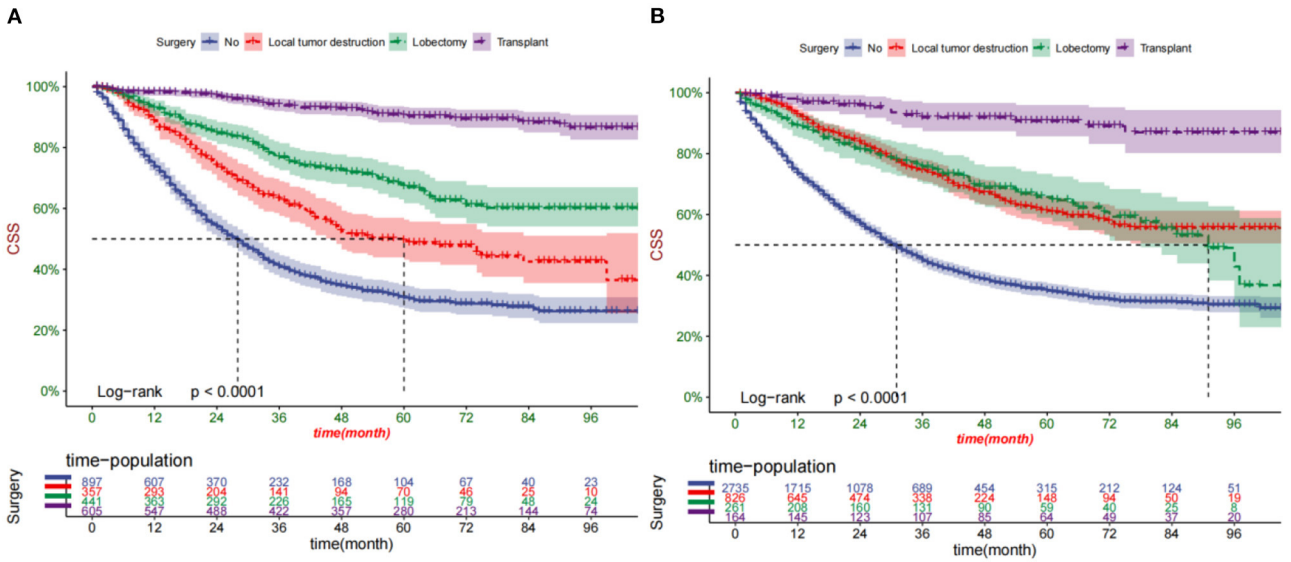


FIGURE 7 | Kaplan-Meier curves of patients with different surgical procedures in the low-risk group **(A)** and high-risk group **(B)**.

overall survival at 1 and 3 years. Still, RFA is resulted in decreased overall survival compared with HR at 5 years (35). A retrospective study showed that lobectomy has better overall survival and higher cancer-specific survival than thermal ablation (36). Moreover, chemotherapy such as sorafenib and lenvatinib is also an effective treatment for HCC (32). Therefore, it is suggested that surgery (OLT, LR, or RFA) combined with chemotherapy is an ideal choice for early-stage HCC patients aged between 40 and 60.

In our prediction model, the AUC of 1-, 3-, 5-year CSS is 0.747 (95%CI: 72.9–76.5), 77.2 (95%CI: 75.4–78.9), and 78.4 (95%CI: 76.2–80.7), respectively, in the training cohort. In the validation cohort, the AUC of 1-, 3-, 5-year CSS is 0.752 (95%CI: 72.3–78.0), 77.5 (95%CI: 75.0–79.9), and 76.6 (95%CI: 73.5–79.7), respectively (**Figure 4**). It indicates that our nomogram has relatively high accuracy in predicting CSS of early-stage HCC patients aged between 40 and 60. In addition, we also drew DCA curves to assess the clinical practicability of the nomogram, which suggested that the prediction model has better clinical application value than the traditional TNM staging system and can more accurately predict the 1-, 3-, 5-year CSS of early-stage HCC patients aged between 40 and 60.

However, our study has limitations. First of all, the nomogram was constructed based on the SEER database. It neither necessarily represents the rest of the world nor includes important factors such as alcohol intake, viral infection, cirrhosis, and liver function. Besides, our study was performed by retrospective analysis, which inevitably has a selection bias that is difficult to adjust. Moreover, our prediction model has only been validated internally, and external validation or further prospective studies are necessary. It is suggested that a multicenter study including more liver function variables be conducted.

CONCLUSION

We explored cancer-specific survival factors in early middle-aged patients with HCC. We found that sex, marriage, race, histological tumor grade, T stage, surgery, chemotherapy, AFP, and tumor size were independent risk factors for patient prognosis. We constructed a new nomogram to predict CSS in patients with early-stage HCC in middle age using these risk factors. This prediction model has good accuracy and reliability, which can help patients and doctors to judge prognosis and clinical decisions.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov/.

AUTHOR CONTRIBUTIONS

JT and CW designed the study, revised the article critically, reviewed, and edited the article. CW, JT, and HL collected and analyzed the data. CW drafted the initial manuscript. All authors approved the final manuscript.

FUNDING

This work was supported by a grant from the Science and Technology Department of Sichuan Province, China (Joint Research Fund, No. 2019YFH0056).

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Petrick JL, McGlynn KA. The changing epidemiology of primary liver cancer. *Curr Epidemiol Rep.* (2019) 6:104–11. doi: 10.1007/s40471-019-00188-3
3. Yang JD, Heimbach JK. New advances in the diagnosis and management of hepatocellular carcinoma. *BMJ.* (2020) 371:m3544. doi: 10.1136/bmj.m3544
4. Peters NA, Javed AA, He J, Wolfgang CL, Weiss MJ. Association of socioeconomics, surgical therapy, and survival of early-stage hepatocellular carcinoma. *J Surg Res.* (2017) 210:253–60. doi: 10.1016/j.jss.2016.11.042
5. Davila JA, El-Serag HB. Racial differences in survival of hepatocellular carcinoma in the United States: a population-based study. *Clin Gastroenterol Hepatol.* (2006) 4:104–10; quiz 4–5. doi: 10.1016/S1542-3565(05)00745-7
6. Ren Z, He S, Fan X, He F, Sang W, Bao Y, et al. Survival prediction model for postoperative hepatocellular carcinoma patients. *Medicine.* (2017) 96:e7902. doi: 10.1097/MD.0000000000007902
7. Sung PS, Yang K, Bae SH, Oh JS, Chun HJ, Nam HC, et al. Reduction of intrahepatic tumour by hepatic arterial infusion chemotherapy prolongs survival in hepatocellular carcinoma. *Anticancer Res.* (2019) 39:3909–16. doi: 10.21873/anticanres.13542
8. Nathan H, Hyder O, Mayo SC, Hirose K, Wolfgang CL, Choti MA, et al. Surgical therapy for early hepatocellular carcinoma in the modern era: a 10-year SEER-medicare analysis. *Ann Surg.* (2013) 258:1022–7. doi: 10.1097/SLA.0b013e31827da749
9. Yamashita YI, Imai K, Yusa T, Nakao Y, Kitano Y, Nakagawa S, et al. Microvascular invasion of single small hepatocellular carcinoma ≤3 cm: predictors and optimal treatments. *Ann Gastroenterol Surg.* (2018) 2:197–203. doi: 10.1002/ags3.12057
10. Li J, Zhou J, Yang PH, Xia Y, Shi YH, Wu D, et al. Nomograms for survival prediction in patients undergoing liver resection for hepatitis B virus-related early-stage hepatocellular carcinoma. *Eur J Cancer.* (2016) 62:86–95. doi: 10.1016/j.ejca.2016.04.011
11. Xiao Z, Yan Y, Zhou Q, Liu H, Huang P, Zhou Q, et al. Development and external validation of prognostic nomograms in hepatocellular carcinoma patients: a population-based study. *Cancer Manag Res.* (2019) 11:2691–708. doi: 10.2147/CMAR.S191287
12. Chen J, Fang A, Chen M, Tuoheti Y, Zhou Z, Xu L, et al. novel inflammation-based nomogram system to predict survival of patients with hepatocellular carcinoma. *Cancer Med.* (2018) 7:5027–35. doi: 10.1002/cam4.1787
13. Yan B, Su BB, Bai DS, Qian JJ, Zhang C, Jin SJ, et al. Practical nomogram and risk stratification system predicting the cancer-specific survival for patients with early hepatocellular carcinoma. *Cancer Med.* (2021) 10:496–506. doi: 10.1002/cam4.3613
14. Shubham S, Jain N, Gupta V, Mohan S, Ariffin MM, Ahmadian A. Identify glomeruli in human kidney tissue images using a deep learning approach. *Soft Comput.* (2021) 1–12. doi: 10.1007/s00500-021-06143-z
15. Movassagh AA, Alzubi JA, Gheisari M, Rahimi M, Mohan S, Abbasi AA, et al. Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model. *J Ambient Intell Human Comput.* (2021) 1–9. doi: 10.1007/s12652-020-02623-6
16. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access.* (2019) 7:81542-81554. doi: 10.1109/ACCESS.2019.2923707
17. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
18. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimed Syst.* (2021). doi: 10.1007/s00530-021-00774-w. [Epub ahead of print].
19. Kumar RL, Khan F, Din S, Band SS, Mosavi A, Ibeke E. Recurrent Neural Network and Reinforcement Learning Model for COVID-19 Prediction. *Front Public Health.* (2021) 9:744100. doi: 10.3389/fpubh.2021.744100
20. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng.* (2021) 18:8444–61. doi: 10.3934/mbe.2021418
21. Kong J, Wang T, Shen S, Zhang Z, Wang W. A nomogram predicting the prognosis of young adult patients diagnosed with hepatocellular carcinoma: a population-based analysis. *PLoS ONE.* (2019) 14:e0219654. doi: 10.1371/journal.pone.0219654
22. Hu C, Yang J, Huang Z, Liu C, Lin Y, Tong Y, et al. Diagnostic and prognostic nomograms for bone metastasis in hepatocellular carcinoma. *BMC Cancer.* (2020) 20:494. doi: 10.1186/s12885-020-06995-y
23. Liu K, Huang G, Chang P, Zhang W, Li T, Dai Z, et al. Construction and validation of a nomogram for predicting cancer-specific

survival in hepatocellular carcinoma patients. *Sci Rep.* (2020) 10:21376. doi: 10.1038/s41598-020-78545-2
24. Ni X, Li D, Dai S, Pan H, Sun H, Ao J, et al. Development and evaluation of nomograms to predict the cancer-specific mortality and overall mortality of patients with hepatocellular carcinoma. *Biomed Res Int.* (2021) 2021:1658403. doi: 10.1155/2021/1658403
25. Lu Z, Sun Z, Liu C, Shi X, Li R, Shao W, et al. Prognostic nomogram for hepatocellular carcinoma with radiofrequency ablation: a retrospective cohort study. *BMC Cancer.* (2021) 21:751. doi: 10.1186/s12885-021-08505-0
26. Song W, Tian C. The effect of marital status on survival of patients with gastrointestinal stromal tumors: a SEER database analysis. *Gastroenterol Res Pract.* (2018) 2018:5740823. doi: 10.1155/2018/5740823
27. Zhang W, Wang X, Huang R, Jin K, Zhang yuan G, Yu W, et al. Prognostic value of marital status on stage at diagnosis in hepatocellular carcinoma. *Sci Rep.* (2017) 7:41695. doi: 10.1038/srep41695
28. Rich NE, Carr C, Yopp AC, Marrero JA, Singal AG. Racial and ethnic disparities in survival among patients with hepatocellular carcinoma in the united states: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol.* (2022) 20:e267–88. doi: 10.1016/j.cgh.2020.12.029
29. Wagle NS, Park S, Washburn D, Ohsfeldt RL, Rich NE, Singal AG, et al. Racial, ethnic, and socioeconomic disparities in curative treatment receipt and survival in hepatocellular carcinoma. *Hepatol Commun.* (2021). doi: 10.1002/hep4.1863. [Epub ahead of print].
30. Goh BK, Teo JY, Chan CY, Lee SY, Jeyaraj P, Cheow PC, et al. Importance of tumor size as a prognostic factor after partial liver resection for solitary hepatocellular carcinoma: implications on the current AJCC staging system. *J Surg Oncol.* (2016) 113:89–93. doi: 10.1002/jso.24099
31. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med.* (1996) 334:693–9. doi: 10.1056/NEJM199603143341104
32. Xie DY, Ren ZG, Zhou J, Fan J, Gao Q. 2019 Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr.* (2020) 9:452–63. doi: 10.21037/hbsn-20-480
33. Rich NE, Parikh ND, Singal AG. Hepatocellular carcinoma and liver transplantation: changing patterns and practices. *Curr Treat Options Gastroenterol.* (2017) 15:296–304. doi: 10.1007/s11938-017-0133-3
34. Carithers RL, Jr. Liver transplantation: will xenotransplantation be the answer to the donor organ shortage? *Trans Am Clin Climatol Assoc.* (2020) 131:270–85.
35. Xu XL, Liu XD, Liang M, Luo BM. Radiofrequency ablation versus hepatic resection for small hepatocellular carcinoma: systematic review of randomized controlled trials with meta-analysis and trial sequential analysis. *Radiology.* (2018) 287:461–72. doi: 10.1148/radiol.2017162756
36. Mills A, Thayer D, Noda C, Salter A, Tao Y, Xing M, et al. Thermal ablation versus surgical resection for localized hepatocellular carcinoma: a population study using the SEER database. *Future Oncol.* (2018) 14:631–45. doi: 10.2217/fon-2017-0447

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wen, Tang and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2022.853294

Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models

C. Kavitha[1]*, *Vinodhini Mani*[1], *S. R. Srividhya*[1], *Osamah Ibrahim Khalaf*[2] and *Carlos Andrés Tavera Romero*[3]

[1] Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India, [2] Al-Nahrain Nanorenewable Energy Research Center, Al-Nahrain University, Baghdad, Iraq, [3] COMBA R&D Laboratory, Faculty of Engineering, Universidad Santiago de Cali, Cali, Colombia

Alzheimer's disease (AD) is the leading cause of dementia in older adults. There is currently a lot of interest in applying machine learning to find out metabolic diseases like Alzheimer's and Diabetes that affect a large population of people around the world. Their incidence rates are increasing at an alarming rate every year. In Alzheimer's disease, the brain is affected by neurodegenerative changes. As our aging population increases, more and more individuals, their families, and healthcare will experience diseases that affect memory and functioning. These effects will be profound on the social, financial, and economic fronts. In its early stages, Alzheimer's disease is hard to predict. A treatment given at an early stage of AD is more effective, and it causes fewer minor damage than a treatment done at a later stage. Several techniques such as Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, and Voting classifiers have been employed to identify the best parameters for Alzheimer's disease prediction. Predictions of Alzheimer's disease are based on Open Access Series of Imaging Studies (OASIS) data, and performance is measured with parameters like Precision, Recall, Accuracy, and F1-score for ML models. The proposed classification scheme can be used by clinicians to make diagnoses of these diseases. It is highly beneficial to lower annual mortality rates of Alzheimer's disease in early diagnosis with these ML algorithms. The proposed work shows better results with the best validation average accuracy of 83% on the test data of AD. This test accuracy score is significantly higher in comparison with existing works.

Keywords: healthcare, prediction, Alzheimer's disease (AD), machine learning, feature selection

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
P. Durgadevi, SRM Institute of Science and Technology, India
Abdul Rehman Javed, Air University, Pakistan
Praveen Kumar, VIT University, India
R. Dhaya, King Khalid University, Saudi Arabia

***Correspondence:**
C. Kavitha
kavitha4cse@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *12 January 2022*
Accepted: *26 January 2022*
Published: *03 March 2022*

Citation:
Kavitha C, Mani V, Srividhya SR, Khalaf OI and Tavera Romero CA (2022) Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models. Front. Public Health 10:853294. doi: 10.3389/fpubh.2022.853294

INTRODUCTION

Alzheimer's Disease (AD) is a progressive neurological condition that leads to short-term memory loss, paranoia, and delusional ideas that are mistaken for the effects of stress or aging. In the United States, this Disease affects about 5.1 million people. AD does not have proper medical treatment. In order to control AD, continuous medication is necessary. AD (1) is chronic so that it can last for years or the rest of your life. Therefore, it is most important to prescribe medication at the appropriate stage so that the brain is not damaged to a great extent. Early detection of this Disease is a tedious and costly process since we must collect a lot of data and use sophisticated tools for prediction and have an experienced doctor involved. Automated systems are more accurate than human assessment and can be used in medical decision support systems because they are not

subject to human errors. Based on previous research on AD, researchers have applied images (MRI scans), biomarkers (chemicals, blood flow), and numerical data extracted from the MRI scans to study this Disease. As such, they were able to determine whether a person was demented or not. In addition to shortening diagnosis time, more human interaction will be reduced by automating Alzheimer's diagnosis. In addition, automation reduces overall costs and provides more accurate results. For example, we can predict whether a patient is demented by analyzing MRI scans and applying prediction techniques. If a person has early-stage Alzheimer's Disease, they are considered demented. By doing so, we can achieve better accuracy.

When a person has Alzheimer's Disease in the early stages, they can usually function without any assistance. In some cases, the person can still work, drive, and partake in social activities. Although this is the case, the person may still feel uneasy or suffer from memory loss, such as not remembering familiar words and locations. People close to the individual notice that they have difficulty remembering their names. By conducting a detailed medical interview, a doctor may identify problems with memory and concentration in the patient. Common challenges in early stage of Alzheimer's Disease include,

- It's hard to remember the right word or name.
- Having difficulty remembering names when meeting new people.
- Working in social settings or the workplace every day can be challenging.
- Having forgotten something that you have just read in a book or something else.
- Having trouble finding or misplacing a valuable object.
- Tasks and activities are becoming increasingly difficult to plan or organize.

Alzheimer's symptoms become more persistent as the Disease progresses. When people suffer from dementia, their ability to communicate, adapt to their environment, and eventually move is lost. It becomes much more difficult for them to communicate pain through words or phrases. Individuals may need substantial assistance with daily activities as their memory, and cognitive skills continue to decline. At this stage, individuals may:

- Personal care and daily activities require 24/7 assistance.
- The consciousness of their surroundings, as well as recent experiences, is lost.
- As you age, you may experience changes in your physical abilities and walking, sitting, and eventually swallowing.
- Communication with others is becoming increasingly difficult.
- Infections, specifically pneumonia, become more prevalent.

Motivation

Under the current conditions, human instinct and standard measurements do not often coincide. In order to solve this problem, we need to leverage innovative approaches such as machine learning, which are computationally intensive and non-traditional. Machine learning techniques are increasingly being used in disease prediction and visualization to offer prescient and customized prescriptions. In addition to improving patients' quality of life, this drift aids physicians in making treatment decisions and health economists in making their analyses. Viewing medical reports may lead radiologists to miss other disease conditions. As a result, it only considers a few causes and conditions. The goal here is to identify the knowledge gaps and potential opportunities associated with ML frameworks and EHR derived data.

Contribution

In our research work, people affected by Alzheimer's Disease are identified and we aims at finding individuals who potentially have Alzheimer's at an early stage. The datasets for Alzheimer's Disease is available on both OASIS and Kaggle which is used for training all patient's data using various machine learning algorithms such as SVM, Random Forest classifier, Decision tree classifier, XGBoost and Voting classifier to effectively distinguish the affected individuals with high degree of efficiency and speed. Finally, an overview of how the Disease has affected the population according to various criteria is analyzed.

Organization

Following are the different sections of our work: Section Related Works address the recent papers on detecting Alzheimer's Disease using Machine learning and Deep learning models. Section Materials and Methods discusses the exploratory data analysis, and different Machine learning classifier models. Results and Discussion section address the performance measures of different Machine Learning models. Finally Section Conclusion concludes the work and discusses the future work.

RELATED WORKS

Alzheimer's Disease is predicted using ML algorithms by using a feature selection and extraction technique, and the classification is conducted based on the oasis longitudinal dataset. The different techniques (2) involved in analyzing brain images for diagnosing diseases of the brain to provide a brief overview. Several major issues are discussed in this article relating to machine learning and deep learning-based brain disease diagnostics based on the results of the reviewed articles. The most accurate method of detecting brain disorders was found in this study and can be used to improve future techniques. Using machine learning and deep learning platforms, this study aims to combine recent research on four brain diseases: Alzheimer's disease, brain tumors, epilepsy, and Parkinson's disease. By using 22 brain disease databases that are used most during the reviews, the authors can determine the most accurate diagnostic method.

Martinez-Murcia et al. (3) uses deep convolutional autoencoders to explore data analysis of AD. Data-driven decomposition of MRI images allows us to extract MRI features that represent an individual's cognitive symptoms as well as the underlying neurodegeneration process. A regression and classification analysis are then performed to examine the distribution of the extracted features in a wide variety of combinations, and the influence of each coordinate of the autoencoder manifold on the brain is calculated. MMSE or

ADAS11 scores, along with imaging-derived markers, can be used for over 80% accuracy to predict AD diagnosis.

A deep neural network is used with layers (4, 5)) that are all connected to perform binary classification. Each hidden layer is activated by a different activation function. A model with the best performance is chosen after k-folds validation Researchers at the Lancet Commission found that about 35% of Alzheimer's risk factors can be modified. The following factors can contribute to these risks: a lack of education, hypertension, obesity, hearing loss, depression, diabetes, lack of physical activity, smoking, and social isolation. Regardless of the impact of these factors at any stage of life, it is beneficial to eliminate them. Studies have suggested (6) that early intervention and treatment of modifiable Alzheimer's risk factors can prevent or delay 30% of cases of Alzheimer's (7). According to the Innovative Midlife Intervention for Alzheimer's Deterrence (In-MINDD) project (8) one way to calculate Alzheimer's risk based on risk factors is by using the Lifestyle for Brain Health (LIBRA) index (9–12). According to the National Academy of Medicine (13, 14) cognitive training, hypertension management, and increased physical activity were the three main categories of dementia intervention. The most common type of Alzheimer's is Alzheimer's Disease (AD). Among the types of Alzheimer's, Vascular Alzheimer's (VaD) is the second most common, followed by Alzheimer's with Lewy bodies. A few other types of Alzheimer's are associated with brain injuries, infections, and alcohol abuse. Tatiq and Barber (15) in their study suggested that Alzheimer's can be prevented by targeting modifiable vascular risk factors because these two types often co-exist in the brain and share some modifiable risk factors. Williams et al. (16) obtained predictions of cognitive functioning based on neuropsychological and demographic data using four different models: SVM, Decision Tree, NN, and Naïve-Bayes. In this case, average values were substituted for the missing values; the accuracy of Naive Bayes was the highest. Data from ADNI study are applied using ten-fold cross-validation (17, 18) and show high correlation between genetic, imaging, biomarker, and neuropsychological outcomes. MRI images from the OASIS dataset (19, 20) are analyzed using voxel-based morphometry. **Table 1** summarizes the recent work on prediction of Alzheimer's disease.

TABLE 1 | Summary of recent work related to AD.

Research study	Dataset	Models	Average classification accuracy
Khan et al. (2)	Inage modality	Machine learning and deep learning models	Study on different ML and DL approaches, and different databases related to brain disease
Saratxaga et al. (21)	OASIS dataset	Deep learning and image processing technique	88%
Sudharsan and Thailambal (22)	ADNI dataset	Machine learning models	75%
Helaly et al. (5)	ADNI dataset	Convolutional neural networks	93% for multiclass AD stages
Shakila Basheer et al. (23)	OASIS dataset	Deep neural networks	92%
Martinez-Murcia et al. (3)	ADNI dataset	Deep learning using convolutional autoencoders	80%
Prajapati et al. (4)	ADNI dataset	Deep neural network binary classifier	85%

TABLE 2 | Dataset description.

S.No.	Attributes	Description
1	ID	Identification
2	M/F	Gender (M if Male, F if Female)
3	Hand	Handedness
4	Age	Age in years
5	EDUC	Years of education
6	SES	Socio Economic Status
7	MMSE	Mini Mental State Examination
8	CDR	Clinical Dementia Rating
9	eTIV	Estimated Total Intracranial Volume
10	nWBV	Normalize Whole Brain Volume
11	ASF	Atlas Scaling Factor
12	Delay	Delay

MATERIALS AND METHODS

The proposed approach consists of three basic steps. Firstly, the Alzheimer's disease dataset (24–26) was loaded into pandas for data preprocessing. This study utilized a longitudinal dataset, so a timeline of the study was necessary to gain further insight into the data. Our first step was to determine how cross-sectional the data appear to be, if either at a baseline or at a particular time. A complete analysis of the data was then conducted, including a comparison of the main study parts and the corresponding data collected during each visit. In this work, longitudinal MRI data is our primary data source. MRI data from 150 patients aged from 60 to 96 were included in the study. We scanned each patient at least once. Everyone is right-handed. Throughout the study, 72 of the patients were classified as "non-demented". At the time of their initial visits, 64 patients were classified as being "Demented," and they remained in this category throughout the study. **Table 2** shows the dataset description of MRI data.

The Machine Learning techniques (26, 27) were applied to Alzheimer's disease datasets to bring a new dimension to predict Disease at an early stage. The raw Alzheimer's disease datasets are inconsistent and redundant, which affects the accuracy of algorithms (28, 29). Before evaluating machine-learning algorithms, data must be effectively prepared for analysis by removing unwanted attributes, missing values, and redundant records. Building a machine-learning model requires splitting the data into training and testing sets. In the following data

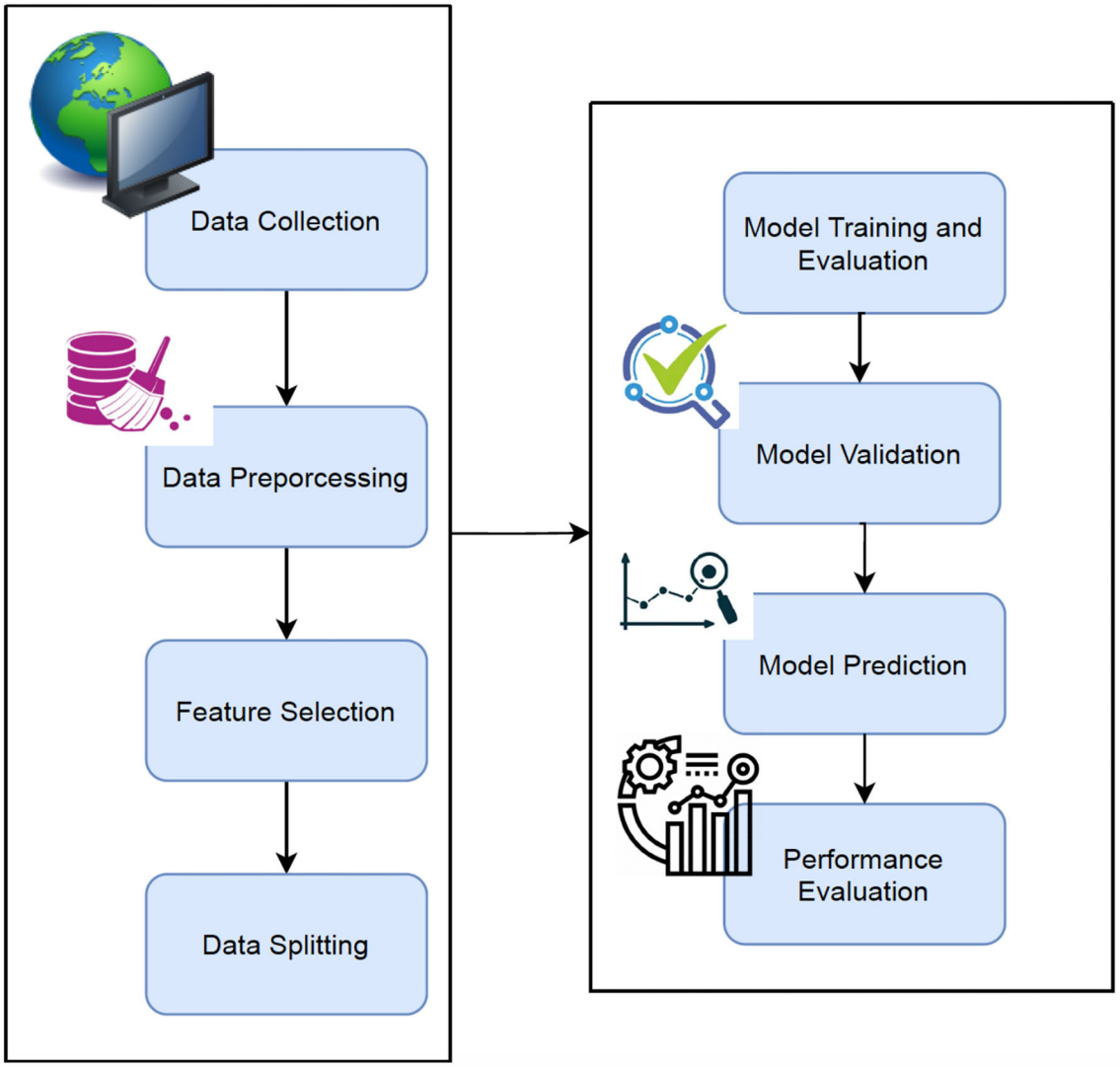


FIGURE 1 | Proposed workflow.

preparation step, the training data were used to create a model, which was then applied to test data to predict Alzheimer's Disease (28, 30, 31). The model was trained from training set data, and test set data were used to test unseen data. Cross-validation was carried out by dividing the dataset into three subsets. Model predictions are made using one subset of the data (test data) and model performance is evaluated using the other subsets (training and validation) of the data. The data had been preprocessed, and we randomly divided it into an 80:20 ratio, with 80% going to training and 20% gone to testing. **Figure 1** describes the system workflow for predicting the Alzheimer's Disease at early stage (32, 33).

Data Preparation

Various data-mining techniques were used to clean and preprocess the data in this phase. As part of this, missing values are handled, features are extracted, and features are transformed, and so on. In the SES column, we identified 9 rows with missing values (34, 35). This issue is addressed in two ways. The simplest solution is to drop the rows with missing values. The other way to fill in missing values is by Imputation (21), which refers to replacing them with their corresponding values. The model should perform better if we impute since we only have 140 measurements. The 9 rows with missing values are removed in the SES attribute and the median value is used for the imputation.

Data Analysis

We have discussed the relationships between each feature of an MRI test and dementia in this section. In order to formulate the relationship of data explicitly using a graph, we conducted this Exploratory Data Analysis process (36, 37) to estimate the correlations before extracting data or analyzing it. The

TABLE 3 | Min, max, and median values of each attribute.

	Min	Max	Mean	Median
EDUC	7	22	14.2	
SES	2	6	2.3	
MMSE	16	30	26.2	
CDR	0	1	0.3	
eTIV	1,120	1,990	1,450	
nWBV	0.55	0.81	0.7	
ASF	0.87	1.43	1.3	

information could be used to interpret the nature of the data later on and to determine what method to use to analyze it. **Table 3** shows the Min, Max and median values of each attribute.

Feature Selection

Feature selection is very important in machine learning. In this work, feature selection is applied to the clinical data of Alzheimer's disease where we have thousands of samples. Feature selection (22) has three methods such as: Filter methods, Wrapper methods, and embedded methods. Filter method is a common method used in the stage of pre-processing. Wrapper methods is another method which core the feature subset. Finally, Embedded method combines the filter and wrapper methods.

The most common and popular feature selection methods are chosen in this work are Correlation coefficient, Information gain, and Chi-Square.

Correlation Coefficient

The covariance between the two variables X and Y is defined as

$$\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

The covariance between two variables measures the linear relationship between them. Using correlation coefficients, it is easy to find a correlation between the various stages of Alzheimer's. The problem with this method is that the data is collected from a broad range of sources, so it becomes very sensitive to outliers.

Information Gain

The entropy of the lower node is subtracted from the entropy of upper node to obtain the Information gain value when the attribute D is selected.

$$Gain\,(D) \;=\; I\,(s1, s2, s3, \ldots\ldots .sn) - E\,(Feature\,D)$$

Chi-Square: Using this method, we can examine categorical variables such as the relationship between food and obesity.

$$Chi - Square \;=\; \frac{(Observed - Expected)^2}{expected}$$

Preparation and Splitting the Data

Figure 2 shows the schematic representation of data splitting stage

1. Select Data: M.F, Age, EDUC, SES, MMSE, eTIV, nWBV, ASF, CDR
2. Train_Data <- round(0.8 * nrow(data)) #Select 80% of train data
3. TrainData_indices <- sample(1:nrow(data), Train_Data). #Vector is created with random indeces
4. TrainML <- data[TrainData_indices,] #Training dataset is generated
5. SplitFormula<- CDR ~ M.F + Age + EDUC + SES + MMSE + eTIV + nWBV
6. N = 5
7. Split<- nWayCrossValidation(nrow(data), n). #5-fold cross validation is generated

Classifier Models

Decision Tree (DT)

An overview of the decision tree gives a tree-based model for dividing the data repeatedly based on the cutoff values of the features. Splitting creates subsets by separating instances into subsets. Intermediate subsets are referred to as internal nodes, while leaves are referred to as leaf nodes. A decision tree is most useful when there is significant interaction between the features and the target.

Random Forest (RF)

A random forest model performs better than a decision tree because it avoids the problem of overfitting. Models based on random forests consist of various decision trees, each slightly different from the others. Using the majority voting algorithm, the ensemble makes predictions based on each individual decision tree model (bagging). As a result, the amount of overfitting is reduced while maintaining the predictive ability of each tree.

Support Vector Machine (SVM)

This method involves determining the class of data points by appropriate hyper planes in a multidimensional space. By using SVM (25), we aim to find a hyperplane that separates cases of two categories of variables that take up neighboring clusters of vectors, one on one side, the other on the other side. Support vectors are those that are closer to the hyperplane. Training and test data are used in SVM. Training data is broken up into target values and attributes. SVM produces a model for predicting target values for test data.

XGBoost

XGBoost stands for eXtreme Gradient BOOSTting. It refers to the process of implementing gradient-boosted decision trees for maximum speed and performance. Due to the sequential nature of model training, gradient boosting machines are generally slow in implementation

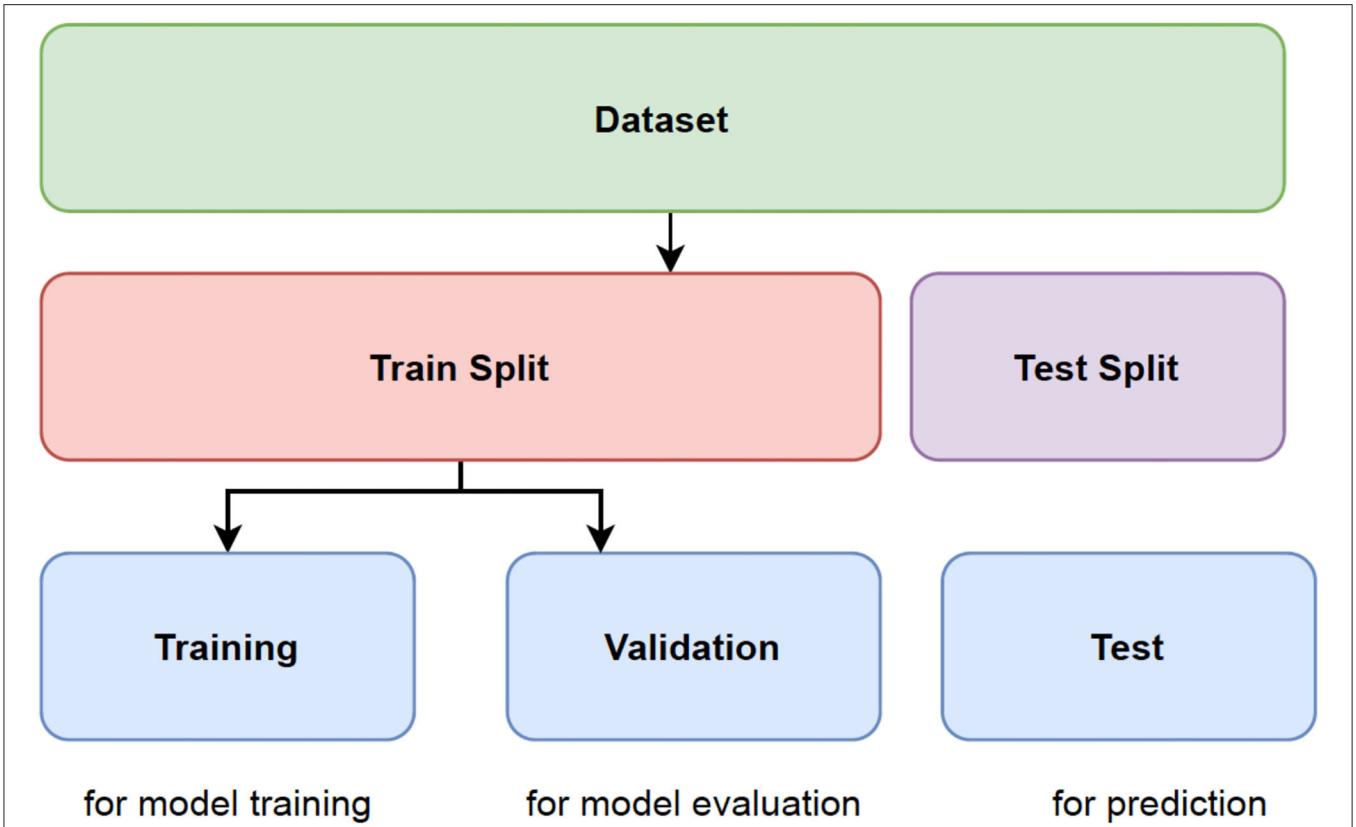


FIGURE 2 | Representation of data splitting.

and not very scalable. XGBoost is focused on speed and performance.

Voting

Voting is one of the simplest ways of combining the predictions from multiple earning algorithms. Voting classifiers aren't actually classifiers but are more like wrappers for multiple ones that are trained and evaluated concurrently in order to benefit from their specific characteristics. We can train data sets using different algorithms and ensembles then to predict the final output. There are two ways to reach a majority vote on a prediction:

Hard voting: The simplest form of majority voting is hard voting. The class with the most votes (Nc) will be chosen in this case. Our prediction is based on the majority vote of each classifier.

Soft voting: This involves adding up the probability vectors for each predicted class (for all classifiers) and choosing the one that represents the highest value (recommended only when the classifiers are well calibrated).

Model Validation

Model validation reduces the overfitting problem. Cross Validation is done to train the ML model and are used to calculate the accuracy of the model. It is a challenging task to make the ML model from noise free. Hence, in this research work, Cross validation is performed which divides the whole dataset into n divisions which is of equal in size. The ML model is trained for every iteration with the n-1 divisions. The performance of the method is analyzed by the mean of all *n*-folds. In this work, the ML model was trained and tested 10 times by applying ten-fold cross validation to the model.

RESULTS AND DISCUSSIONS

We evaluate various performance metrics like accuracy, precision, recall and F1 score. To determine the best parameters for each model, we perform 5-fold cross-validation: Decision Tree, SVM, Random Forests, XGBoost and Voting. Finally, we compare accuracy of each model. Several metrics and techniques were used to identify overfitting and parameter tuning issues after the models were developed. Performance evaluations can either be binary or multiclass and are described using the confusion matrix. A learning model was developed to distinguish true Alzheimer's disease affected people from a given population and a novel Machine Learning classifier was developed and validated to predict and separate true Alzheimer's disease affected people. The following evaluation measures were calculated using these components: precision, recall, accuracy,

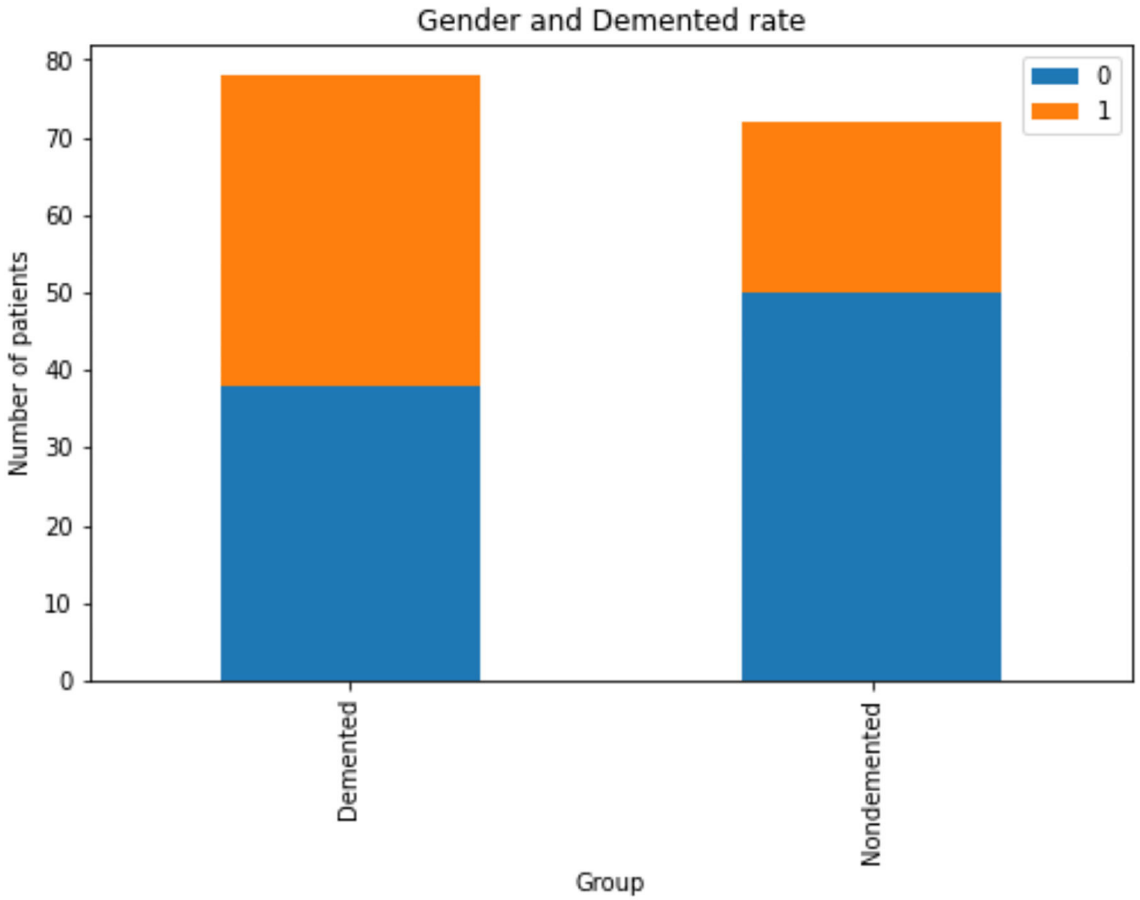


FIGURE 3 | Analysis of demented and non-demented rate based on gender, Gender group Female = 0, Male = 1.

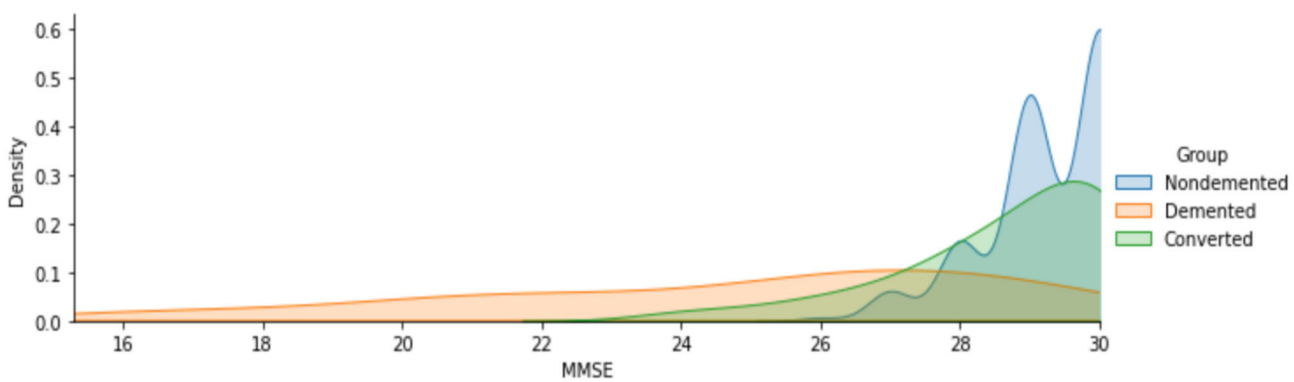


FIGURE 4 | Analysis of MMSE scores for demented and non-demented group of patients.

and F-score. Based on this study, recall (sensitivity) is the proportion of people accurately classified as having Alzheimer's. The precision of Alzheimer's diagnosis is the rate of people correctly classified as not having the disease. Alternatively, F1 represents the weighted average of recall and precision, while accuracy represents the proportion of people correctly classified. According to the results, the patient receives a report that tells him or her what stage of Alzheimer's Disease he or she is currently in. It is very important to detect the stages because the stages are based on the responses of the patients. In addition, knowing the stage helps doctors better understand how the Disease is affecting them. This research used these environments, tools, and libraries to conduct its experiments and analysis:

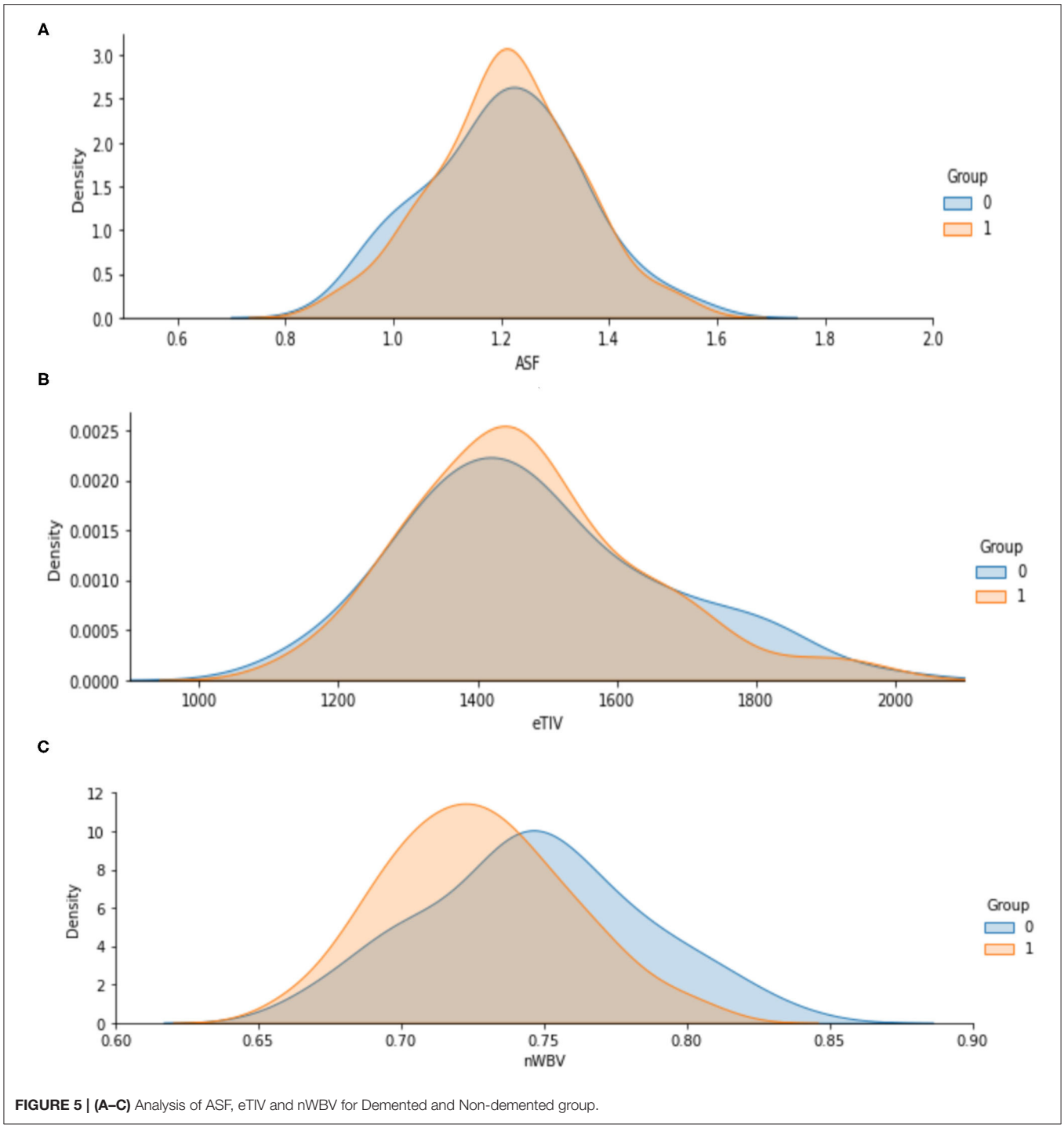


FIGURE 5 | (A–C) Analysis of ASF, eTIV and nWBV for Demented and Non-demented group.

a) Environments Used: Python 3
b) Scikit-learn libraries for machine learning

The **Figure 3** indicates that men are more likely than women to have dementia. **Figure 4** that the non-demented group had much higher MMSE (Mini-Mental State Examination) scores than those with dementia.

The **Figures 5A–C** shows the analyzed value of ASF, eTIV and nWBV for Demented and Non-demented group of people. As indicated by the graph in **Figure 5**, the Non-demented group has a higher brain volume ratio than the Demented group. The reason for this is that the diseases influence the brain tissues causing them to shrink. **Figure 6** shows the analyzed results of EDUC for Demented and Non-demented people.

Figure 7 shows the analysis on age attribute to find the percentage of people affected based on the demented and non-demented group. It is observed that a higher percentage of Demented patients are 70-80 years old than non-demented

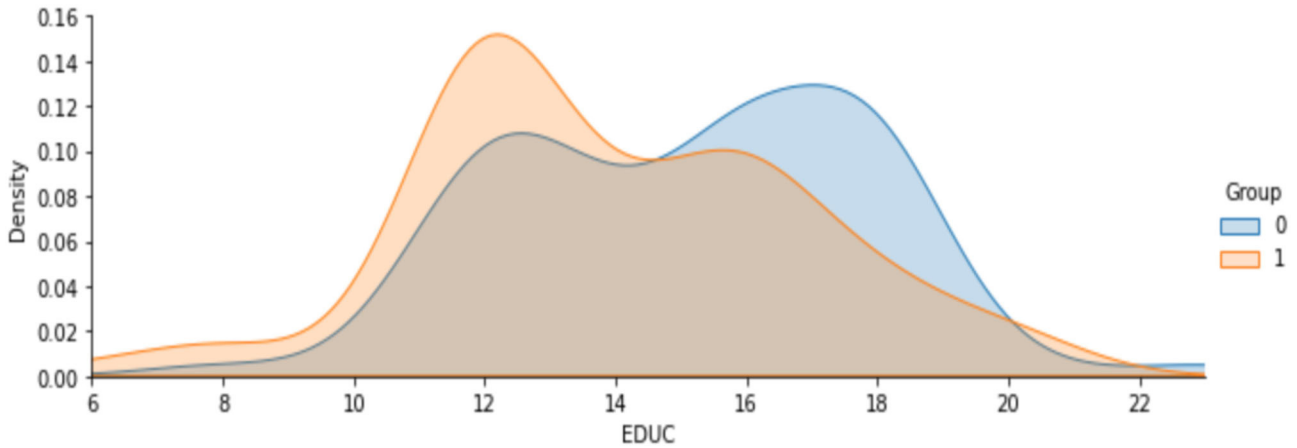


FIGURE 6 | Analysis on years of education.

FIGURE 7 | Analysis on people affected by demented and non-demented group based on age.

TABLE 4 | Performance comparison of different ML models.

Model	Accuracy	Precision	Recall	F1-score
Decision tree classifier	80.46%	0.80	0.79	0.78
Random forest classifier	86.92%	0.85	0.81	0.80
Support vector machine	81.67%	0.77	0.70	0.79
XGBoost	85.92%	0.85	0.83	0.85
Voting classifier	85.12%	0.83	0.83	0.85

patients. It is likely that people with that kind of Disease have a low survival rate. Only a few people are over 90 years old.

From the above all analysis on the attributes, the following are the summary on intermediate results.

1. It is more likely for men to have demented, or Alzheimer's Disease, than for women.
2. In terms of years of education, demented patients were less educated.
3. Brain volume in non-demented groups is greater than in demented groups.
4. Among the demented group there is a higher concentration of 70-80-year-olds than in the non-demented patients.

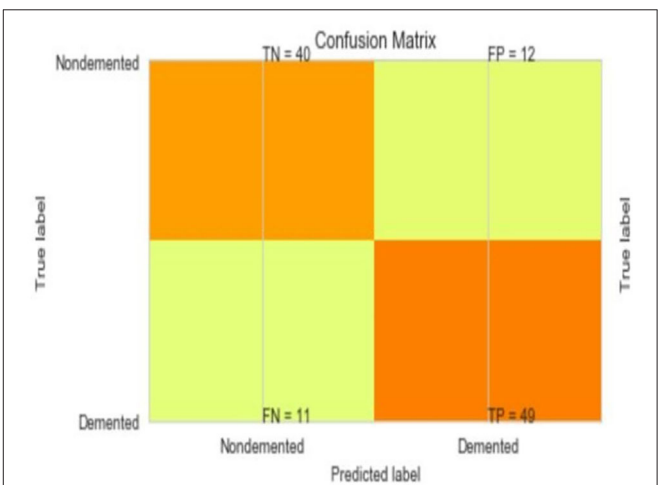


FIGURE 8 | Confusion matrix for decision tree.

Table 4 shows the performance comparison of accuracy, precision, recall, and F1 score for different ML models. The performance measures are defined as,

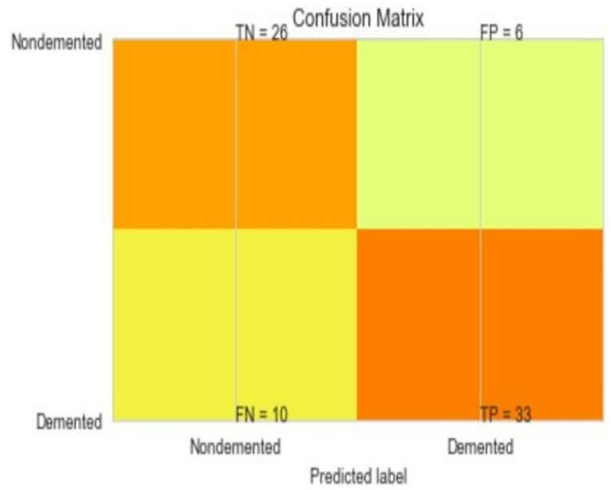


FIGURE 9 | Confusion matrix for random forest.

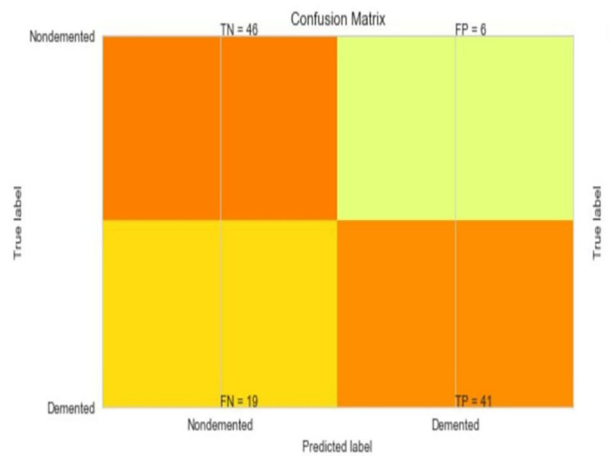


FIGURE 10 | Confusion matrix for SVM.

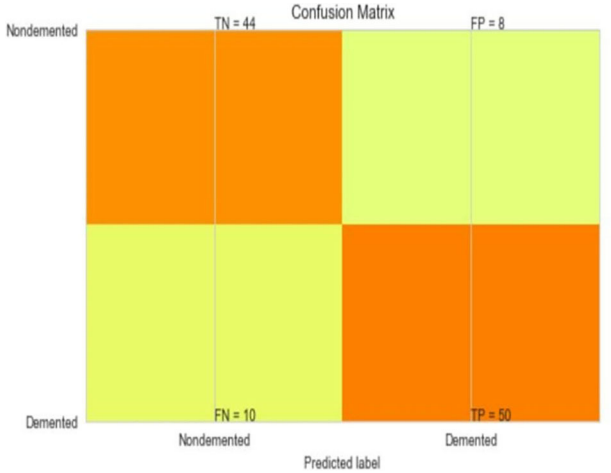


FIGURE 11 | Confusion matrix for XGBoost.

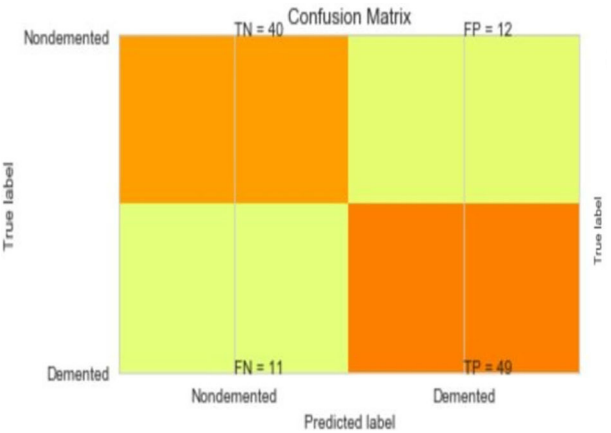


FIGURE 12 | Confusion matrix for soft voting classifier.

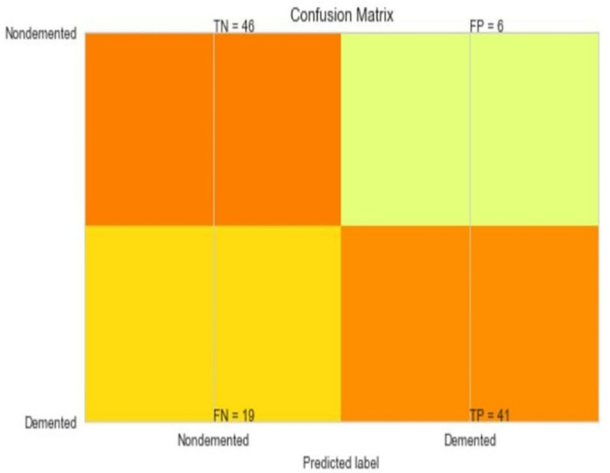


FIGURE 13 | Confusion matrix for hard voting classifier.

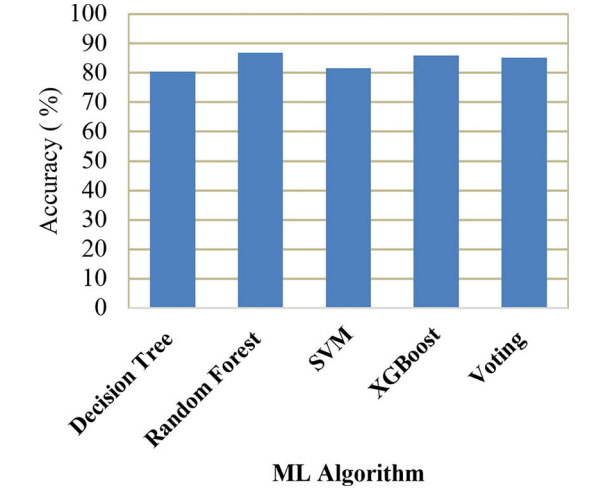


FIGURE 14 | Comparison of accuracy.

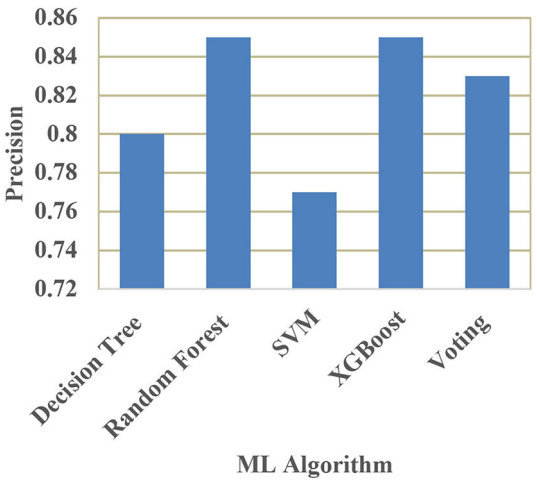


FIGURE 15 | Comparison of precision.

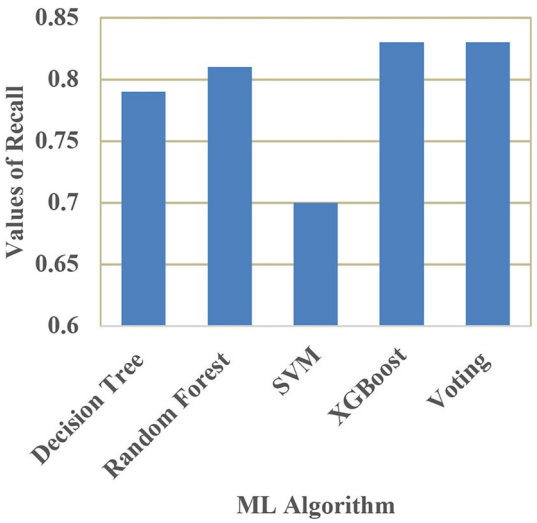


FIGURE 16 | Comparison of recall.

Accuracy: It is the measure of finding the proportion of correctly classified result from the total instances.

$$Accuracy(inPercentage) = \frac{TN + TN}{TP + TN + FP + FN} \times 100$$

Precision: This measures the number of correctly predicted positive rate divided by the total predicted positive rates. If the Precision value is 1, it is meant as a good classifier.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall is a true positive rate. If the recall is 1, it is meant as a good classifier.

$$Recall(inPercentage) = \frac{TP}{TP + FN}$$

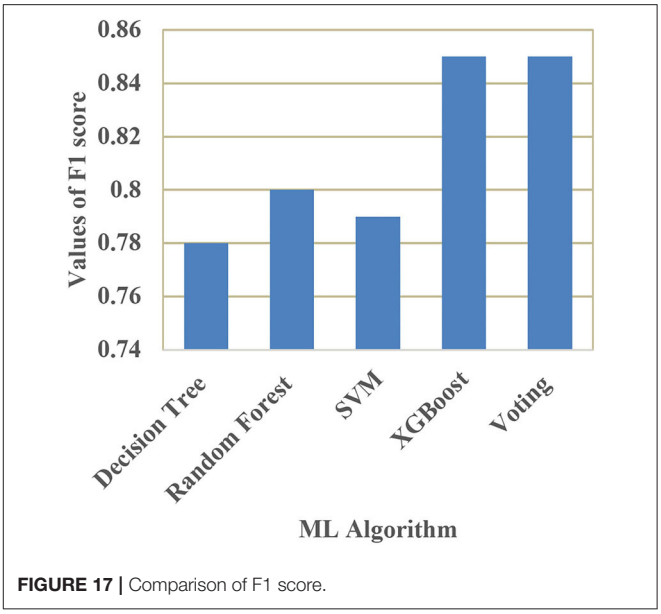


FIGURE 17 | Comparison of F1 score.

F1 Score: It is a measure which considers both Recall and Precision parameters. F1 score becomes 1 only when both the measure such as Recall and Precision is 1.

$$F1Score(inPercentage) = 2^* \frac{Recall^* Precision}{Recall + Precision}$$

The most common metrics are the conversions of the True Positive (TP), the False Positive (FP), the True Negative (TN), and the False Negative (FN) metrics. **Figures 8–13** shows the confusion matrix for Decision tree, Random Forest, SVM, XG boost, Soft, and Hard Voting classifier ML models.

A comparison of training and testing accuracy has been conducted for each model to eliminate overfitting. For each model, precision, recall, accuracy, and F1-score are shown in **Table 3**. Based on the analysis showed in the **Table 3**, the results approved that the best and ideal techniques, which have a good performance, are random forest, and XGBoost. The accuracy value of Voting classifier model is also closer to the random forest, and XGBoost models. All the experimental results (the average accuracy, precision, recall, and F measure of each model) were collected for extra analysis. The comparative analyses among all the Machine Learning models in terms of accuracy, precision, recall, and F1 score are presented graphically in **Figures 14–17** respectively.

CONCLUSIONS

Alzheimer's is a major health concern, and rather than offering a cure, it is more important to reduce risk, provide early intervention, and diagnose symptoms early and accurately. As seen in the literature survey there have been a lot of efforts made to detect Alzheimer's Disease with different machine learning algorithms and micro-simulation methods; however, it remains a challenging task to identify relevant attributes that can

detect Alzheimer's very early. The future work will focus on the extraction and analysis of new features that will be more likely to aid in the detection of Alzheimer's Disease, and on eliminating redundant and irrelevant features from existing feature sets to improve the accuracy of detection techniques. By adding metrics like MMSE and Education to our model, we'll be able to train it to distinguish between healthy adults and those with Alzheimer's.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://www.kaggle.com/jboysen/mri-and-alzheimers?select=oasis_cross-sectional.csv.

AUTHOR CONTRIBUTIONS

CK: research concept and methodology and writing—original draft preparation. VM: review and editing. SS: supervision. OK and CT: validation. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been funded by Dirección General de Investigaciones of Universidad Santiago de Cali under call No. 01-2021.

REFERENCES

1. Sivakani GA, Ansari R. Machine learning framework for implementing Alzheimer's disease. *Int Conferen Commun Signal Process.* (2020) 12:588–92. doi: 10.1109/ICCSP48568.2020.9182220
2. Khan P, Kader MF, Islam SR, Rahman AB, Kamal MS, Toha MU, et al. Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. *IEEE Access.* (2021) 9:37622–55. doi: 10.1109/ACCESS.2021.3062484
3. Martinez-Murcia FJ, Ortiz A, Gorriz JM, Ramirez J, Castillo-Barnes D. Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. *IEEE J Biomed Health Inform.* (2020) 24:17–26. doi: 10.1109/JBHI.2019.2914970
4. Prajapati R, Khatri U, Kwon GR. "An efficient deep neural network binary classifier for alzheimer's disease classification," In: *International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. (2021), p. 231–234.
5. Helaly HA, Badawy M, Haikal AY. Deep learning approach for early detection of Alzheimer's disease. *Cogn Computing.* (2021) 21:1–17. doi: 10.1007/s12559-021-09946-2
6. Yaffe K. Modifiable risk factors and prevention of dementia: what is the latest evidence. *JAMA Intern Med.* (2018) 178:281–2. doi: 10.1001/jamainternmed.2017.7299
7. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley, D, et al. Dementia prevention, intervention, and care. *The Lancet.* (2017) 390:2673–73. doi: 10.1016/S0140-6736<17>31363-6
8. O'Donnell CA, Manera V, Köhler S, Irving K. Promoting modifiable risk factors for dementia: is there a role for general practice? *British J General Pract.* (2015) 65:567–8. doi: 10.3399/bjgp15X687241
9. Sulaiman N, Abdulsahib G, Khalaf O, Mohammed MN. "Effect of Using Different Propagations of OLSR and DSDV Routing Protocols", In *Proceedings of the IEEE International Conference on Intelligent Systems Structureing and Simulation*. (2014), pp. 540-5.
10. Deckers K, van Boxtel MP, Schiepers OJ, de Vugt M, Muñoz Sánchez JL, Anstey KJ, et al. Target risk factors for dementia prevention: a systematic review and Delphi consensus study on the evidence from observational studies. *Int J Geriatric Psychiatry.* (2015) 30:234–46. doi: 10.1002/gps.4245
11. Schiepers OJ, Köhler,S., Deckers K, Irving K, O'donnell CA, Van den Akker, et al. Lifestyle for Brain Health (LIBRA): a new model for dementia prevention. *Int J Geriatric Psychiatry.* (2018) 33:167–75. doi: 10.1002/gps.4700
12. Vos SJ, Van Boxtel MP, Schiepers OJ, Deckers K, De Vugt M, Carrière I, et al. Modifiable risk factors for prevention of dementia in midlife, late life and the oldest-old: validation of the LIBRA Index. *J Alzheimer's Dis.* (2017) 58:537–47. doi: 10.3233/JAD-161208
13. Osamh Khalaf I, Ghaida M, Abdulsahib D. Energy efficient routing and reliable data transmission protocol in WSN. *Int J Adv Soft Comput Applicat.* (2020) 12:45–53.
14. National Academies of Sciences, Engineering, and Medicine. *Preventing cognitive decline and dementia: A way forward*. London: The National Academies Press (2018).
15. Tariq S, Barber PA. Dementia risk and prevention by targeting modifiable vascular risk factors. *J Neurochemistr.* (2018) 144:565–81. doi: 10.1111/jnc.14132
16. Williams., Jennifer A, Weakley A, Cook MS, Edgecombe DJ. "Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia," *In Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*. (2018), pp. 71–6.
17. Khalaf OI, Sabbar BM. A modified algorithm for improving lifetime WSN. *J Eng Appl Sci.* (2018) 13:9277–82.
18. Khalaf OI, Abdulsahib GM, Sabbar BM. Optimization of wireless sensor network coverage using the Bee Algorithm. *J Inf Sci Eng.* (2020) 36:377–86.
19. Chi CL, Oh, Borson WS. "Feasibility Study of a Machine Learning Approach to Predict Dementia Progression," in *International Conference: In Health care Informatics (ICHI)*. (2015), p. 450.
20. Chyzhyk A, Savio D. *Feature extraction from structural MRI images based on VBM: data from OASIS database, University of The Basque Country, Internal research publication* (2010).
21. Saratxaga CL, Moya I, Picón A, Acosta M, Moreno-Fernandez-de-Leceta A, Garrote E, et al. MRIDeep learning-based solution forAlzheimer's Disease Prediction. *J.Pers. Med.* (2021) 11:902. doi: 10.3390/jpm11090902
22. Sudharsan M, Thailambal G. *Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA), Materials Today: Proceedings* (2021).
23. Basheer S, Bhatia S, Sakri SB. "Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset," in *IEEE Access*. (2021) 9:42449–42462. doi: 10.1109/ACCESS.2021.3066213
24. Khalaf OI, Abdulsahib GM. Frequency estimation by the method of minimum mean squared error and P-value distributed in the wireless sensor network. *J Informat Sci Eng.* (2019) 35:1099–112.
25. Ogudo KA, Muwawa J, Ibrahim Khalaf O, Daei Kasmaei H. A device performance and data analytics concept for smartphones' IoT services and machine-type communication in cellular networks. *Symmetry.* (2019) 11:593–609. doi: 10.3390/sym11040593
26. Reddy GT, Reddy MP, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intel.* (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
27. Abdulsahib GM, Khalaf OI. Accurate and effective data collection with minimum energy path selection in wireless sensor networks using mobile sinks. *J Informat Technol Manage.* (2020) 13:139–53.
28. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PK, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Ambient Intell Human Computing.* (2020) 21:1–4. doi: 10.1007/s12652-020-01963-7
29. Salman AD, Khalaf OI, Abdulsahib GM. An adaptive intelligent alarm system for wireless sensor network. *Indonesian J Electric Eng Comput Sci.* (2019) 15:142–7. doi: 10.11591/ijeecs.v15.i1.pp142-147
30. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic Deep Q Network for real-time path planning in censorious robotic procedures using force sensors. *J Real-Time Image Proc.* (2021) 18:773–1785. doi: 10.1007/s11554-021-01122-x

31. Khalaf OI, Abdulsahib GM, Kasmaei HD, Ogudo KA. A new algorithm on application of blockchain technology in live stream video transmissions and telecommunications. *Int J e-Collaboration (IJeC).* (2020) 16:16–32. doi: 10.4018/IJeC.2020010102
32. Abdulsahib GM, Khalaf OI. An improved algorithm to fire detection in forest by using wireless sensor networks. *Int J Civil Eng Technol (IJCIET).* (2018) 9:369–77.
33. Abdulsahib GM, Khalaf IO. Comparison and evaluation of cloud processing models in cloud-based networks. *Int J Simul Syst Sci Technol.* (2018) 19:5. doi: 10.5013/IJSSST.a.19.05.26
34. Khalaf OI, Abdulsahib GM. Optimized dynamic storage of data (ODSD) in IoT based on blockchain for wireless sensor networks. *Peer-to-Peer Netw.* (2021) 21:255. doi: 10.1007/s12083-021-01115-4
35. Javed AR, Fahad LG, Farhan AA, Abbas S, Srivastava G, Parizi RM, et al. Automated cognitive health assessment in smart homes using machine learning. *Sustain Cities Soc.* (2020) 65:102572. doi: 10.1016/j.scs.2020.102572
36. Maddikunta, PR, Gadekallu, TR, Iwendi, C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021) 21:1–12. doi: 10.1049/ipr2.12222
37. Javed AR, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. PP-SPA: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Process Lett.* (2021) 21:1–8. doi: 10.1007/s11063-020-10414-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kavitha, Mani, Srividhya, Khalaf and Tavera Romero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2022.849185

A Multi-Agent Formalism Based on Contextual Defeasible Logic for Healthcare Systems

Salwa Muhammad Akhtar[1], Makia Nazir[1], Kiran Saleem[2], Rana Zeeshan Ahmad[3], Abdul Rehman Javed[4], Shahab S. Band[5]* and Amir Mosavi[6,7,8]**

[1] Computer Science Department, Faculty of Computer Science and IT, University of Lahore, Lahore, Pakistan, [2] School of Software, Dalian University of Technology, Dalian, China, [3] Department of Information Technology, University of Sialkot, Sialkot, Pakistan, [4] Department of Cyber Security, Air University, Islamabad, Pakistan, [5] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliou, Taiwan, [6] Institute of Information Society, University of Public Service, Budapest, Hungary, [7] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [8] Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Randhir Kumar,
Indian Institute of Technology Hyderabad, India
Weizheng Wang,
City University of Hong Kong, Hong Kong SAR, China
Praveen Kumar,
VIT University, India

***Correspondence:**
Abdul Rehman Javed
abdulrehman.cs@au.edu.pk
Shahab S. Band
shamshirbands@yuntech.edu.tw
Amir Mosavi
amir.mosavi@nik.uni-obuda.hu

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 05 January 2022
Accepted: 24 January 2022
Published: 03 March 2022

Citation:
Akhtar SM, Nazir M, Saleem K, Ahmad RZ, Javed AR, S. Band S and Mosavi A (2022) A Multi-Agent Formalism Based on Contextual Defeasible Logic for Healthcare Systems. Front. Public Health 10:849185. doi: 10.3389/fpubh.2022.849185

In the last decade, smart computing has garnered much attention, particularly in ubiquitous environments, thus increasing the ease of everyday human life. Users can dynamically interact with the systems using different modalities in a smart computing environment. The literature discussed multiple mechanisms to enhance the modalities for communication using different knowledge sources. Among others, Multi-context System (MCS) has been proven quite significant to interlink various context domains dynamically to a distributed environment. MCS is a collection of different contexts (independent knowledge sources), and every context contains its own set of defined rules and facts and inference systems. These contexts are interlinked *via* bridge rules. However, the interaction among knowledge sources could have the consequences such as bringing out inconsistent results. These issues may report situations such as the system being unable to reach a conclusion or communication in different contexts becoming asynchronous. There is a need for a suitable framework to resolve inconsistencies. In this article, we provide a framework based on contextual defeasible reasoning and a formalism of multi-agent environment is to handle the issue of inconsistent information in MCS. Additionally, in this work, a prototypal simulation is designed using a simulation tool called NetLogo, and a formalism about a Parkinson's disease patient's case study is also developed. Both of these show the validity of the framework.

Keywords: healthcare system, NetLogo, web ontology, multi-agent system, medical internet of things (IoT)

1. INTRODUCTION

Ubiquitous computing deals with invisibly interweaving the real world with various agents embedded seamlessly in everyday objects of lives and connected through dedicated networks and media to make our everyday life easier and more efficient (1–3). When the agents were initially designed, they were used solely to detect user location, however, with technological advancements, they gradually were able to adapt to interfaces, increase the precision of information retrieval, discover services users may require, and most importantly build smart environments (4). Smart environments are the physical environments used in daily human life that are seamlessly embedded with tiny smart devices equipped with sensors, actuators, and computational elements (agents).

These physically embedded devices are connected through a continuous data collection and processing network to enable various pervasive applications, services, and smart environments to perform efficiently. The usage of various connected agents for data retrieval, processing, and exchanging make up a system called Multi-Agent System (MAS) (5).

A multi-agent system is a computerized system comprising multiple interacting intelligent agents (6). These intelligent agents can collect data, process it, and then send it to the controller/central system for further analysis. While this level of intelligence and functionality is more than enough for some applications, for others, this is not. These are the applications or systems that needed information from multiple resources (domains) to execute their required operations (4). This problem leads to a solution termed Multi-Context Systems (MCS).

Equilibrium is defined as a belief state that comprises a belief set, consisting of beliefs of each context in the MCS, such that an MCS is said to be in equilibrium if all the heterogeneous contexts have consistent beliefs (7). Inconsistency is a significant concern depriving MCS of reaching its full potential. Inconsistency makes it impossible for the system to conclude from the provided data. Suppose a system cannot derive a decision, then it is considered a failure. This work tries to ensure that the system maintains its equilibria state. This is the state in which each element of the system has a single belief, i.e., the decision should be the same according to each element of the system. The equilibria are lost whenever the system receives contradictory or inconsistent information. This makes the overall system not provide a conclusion, causing it to become purposeless. This becomes a severe issue, especially in the safety-critical system dealing with life and death situations in healthcare systems (8).

The proposed system uses a case study about a Parkinsons' disease patient to show its correctness. The system includes a Multi-Context system (MCS) consisting of several heterogeneous contexts. Several agents acquire information, process data, and perform reasoning accordingly in each context in MCS. As physical implementation of such a system is not feasible due to financial and technological constraints, virtual agents model the proposed smart environment. It is assumed that these virtual agents are intelligent; thus, they are intelligent virtual agents. Virtual agents can also be considered a type of software that collects, processes, and shares data just like a physical sensor and agent. It is assumed that data is acquired from virtual sensors for data acquisition. In the proposed framework, intelligent agents acquire contextual information from the sensors and perform reasoning accordingly. For each domain, an ontology is developed using Protégé, a free, open-source ontology editor and management system. Each of these ontologies is connected using mapping rules. These ontologies are mapped on their corresponding description logic. These description logic-based ontologies are then linked together using distributed description logic (DDL) formalism. Once the system gets settled, the flow of information begins due to the communication between the heterogeneous contexts. After this, the system's conflicting information is dealt with using CDL (9). Contextual defeasible reasoning can handle inconsistencies occurring due to contradictory information (8).

The rest of this article is structured as follows. Section Related Work presents the related work. Section Contextual Defeasible Reasoning (CDL) Based Multi-Agent Formalism presents the extensive description of the architecture of the framework proposed and the inclusion of Contextual Defeasible Reasoning in it. Section Framework Development discusses the contexts developed for the framework and their corresponding ontologies. Section Temporal Logic Formalism covers the formalism of the proposed work using temporal logic. Section Simulation of Proposed System discusses the simulation of the system done using NetLogo. In section Algorithm the proposed algorithm for the system is discussed in detail. Finally, this work is discussed overall and concluded in sections Limitations and Future Scope and Conclusion, respectively.

2. RELATED WORK

Authors in (9) proposed a methodology for handling inconsistency produced in a Multi-Context System (MCS) by utilizing the consistency-based and abduction-based techniques. In the first approach, pairs of sets of mapping rules are considered, such that deactivating the rules in one set and forcing the rules in the other to be active allows to re-establish consistency. While in the second approach, a pair of rules were considered whose joint (de-)activation reproduced the observed inconsistency. In (8), the authors proposed a framework based on Contextual Defeasible Logic (CDL) for distributed MCS. The authors deployed the Meta-rules technique to perform reasoning on the system-acquired data.

In (10), the authors present a smart city ontology called KM4City (Knowledge Model for City). This ontology includes data coming from various heterogeneous sources and mapped onto it. Since ontologies are flexible, therefore, the addition of new information can easily be managed. Static and dynamic data from various heterogeneous sources is obtained and stored in RDF syntax, accessed using SPARQL queries. This data was from public and private sources, such as people sending information from their devices. The primary purpose of the author's KM4City is to help citizens on roads, mainly those stuck in traffic, and especially helping the ambulances find an optimal path to their destinations, be it the hospital or the location of a patient. The proposed work, according to the authors, overcomes the gap of combining public and private data from various heterogeneous sources. Various applications like those of public administrations and enterprises can be made using this framework. The information collected for the scope of the proposed paper was limited to road graphs, services available on the roads, and traffic sensors.

In (11), the authors have used an RPI-CAM-V2 pi camera, soil moisture sensor, and a temperature sensor and connected them to a Raspberry pi microcontroller. A raspberry pi is a circuit board that can be attached to a computer monitor through the USB port. This system has especially been developed for the farmers or the caretakers of agricultural lands to increase the efficiency of the agricultural systems by automating the irrigation process. It then sends the data to the Thingspeak mobile application (an android application available on the play

TABLE 1 | Analysis of related work.

Reference	Technique	Domain
(8)	Defeasible logic theory using the notion of meta-rules	Multi-context distributed systems
(10)	Ontology-driven modeling	Smart city
(14)	Machine learning algorithms	Smart healthcare
(9)	Consistency-based and abduction-based techniques	Multi-Context System (MCS)
(12)	Industry 5.0 supporting technologies	Internet of Everything (IoE)
(13)	Grid and place neuron model	2D virtual environment
(11)	Raspberry-pi circuit board	Smart agriculture
(16)	Internet of Medical Things (IoMT)	Smart healthcare
(15)	Blockchain technique	Internet of Things (IoT)
(17)	Elliptic Curve Digital Signature Algorithm (ECDSA)	Large-scale batch verification

store of all android devices), where the farmers can easily have access to the environmental conditions of the fields without being in the area. The authors have also used the Telegram mobile application for fetching data from raspberry pi when in remote locations. Telegram application is a messaging application similar to Whatsapp, except for creating bots that interact with both humans and machines.

In (12), the authors have presented a survey on the application of technologies of industry 5.0 in different domains. Industry 5.0 is mainly used to increase the efficiency and effectiveness of large-scale product manufacturing. However, like any other technology, this also comes with many drawbacks, mainly privacy and security dilemmas. In (13), the authors have presented a grid and place neuron model in cognitive tasks applications. The 2D virtual environment was created to deploy this system, which resulted in 92.27 % localization accuracy.

In (14), the authors have highlighted the usage of ML algorithms for the diagnosis, analysis, and prediction of stroke. Moreover, the authors used the Antlion optimization (ALO) algorithm using a deep learning model in minimal time consumption for optimal hyper-parameters selection. The training time for the proposed model is 38 % as a positive outcome considered for the experimental superiority results.

In (15), the authors have proposed a novel pairing-free certificateless scheme using blockchain technique and a CLS scheme using a smart contract. After this, they simulated the Type-I and Type-II adversaries to verify their scheme. After a thorough analysis, the authors deduced that their work reduced 40.0% computation cost and 94.7% communication cost. In (16), the authors have highlighted the emergence of heterogeneous Internet of Medical Things (IoMT) (e.g., for smart healthcare systems) that is used for sending huge bulk of patient's data for disease analysis to central cloud servers. However, it is prone to many security issues that can be overcome using AKA p. The authors have proposed an authentication protocol using blockchain technology and physically unclonable functions (PUF) in their work. In addition, biometric information is dealt with using a fuzzy extractor scheme. Their analysis proved that their work requires the most negligible computation and communication cost among the compared schemes.

This work is novel in the way that it proposes a smart decision-making healthcare system based on formalism technique that can help the medical professionals in their everyday routine and the patients living in far-off areas. The modeling of the system was done using ontologies so a clear picture of the system and the flow of information in it can be obtained. In order to check the validity of this work, a simulation was created using the NetLogo simulation tool. The existing approaches are summarized in **Table 1**. This work is different from others such that it uses contextual defeasible logic to resolve inconsistencies in multi-context systems. This technique is helpful when the entire system needs to be controlled automatically with the help of rule-based formalism in order to allow the system to reach a conclusion about the patient's health and provide the medical authority about the current situation.

3. CONTEXTUAL DEFEASIBLE REASONING (CDL) BASED MULTI-AGENT FORMALISM

This section covers the CDL based on multi-agent formalism. In the context-aware MAS, every agent has a set of defeasible rules in their knowledge base where the contextual information and reasoning strategy is stored. Multi-context system notions are used for modeling heterogeneous systems (18–20). Multiple ontologies can send contextual information (e.g., rules, and facts). Each knowledge source sends the contextual information to its corresponding agents. After that, agents perform their reasoning mechanism based on the pre-defined rules. Reasoning types are two called, locally and globally. In local reasoning, the rules are performed by obtaining, through a single ontology. In distributed reasoning, multiple ontologies extract the sets of rules and facts by which each agent performs its reasoning. A framework of contextual defeasible reasoning based on heterogeneous formalism is proposed in this thesis. The framework uses the data acquired by sensors and analyzed and shared by agents, working in various heterogeneous contexts/domains with their respective systems. The following case study has been developed, which will form the basis of framework modeling and development.

Figure 1 shows the working of the proposed system. First, the virtual sensors continuously send data to a database. This database contains additional information about the patient, his allergies, and insurance. The information from the database (which also contains data from the sensors) is sent to the proposed system, which is the intelligent system responsible for making a decision. The intelligent system responsible for making a decision acquires this information and tries to derive a conclusion. In case of no inconsistency during conclusion making, then the system prescribes the treatment for the disease and keeps a record of the severity level of the patient. But if an inconsistency does occur, then the system first handles this inconsistency by using contextual defeasible logic and then prescribes treatment and severity level. After allotting severity

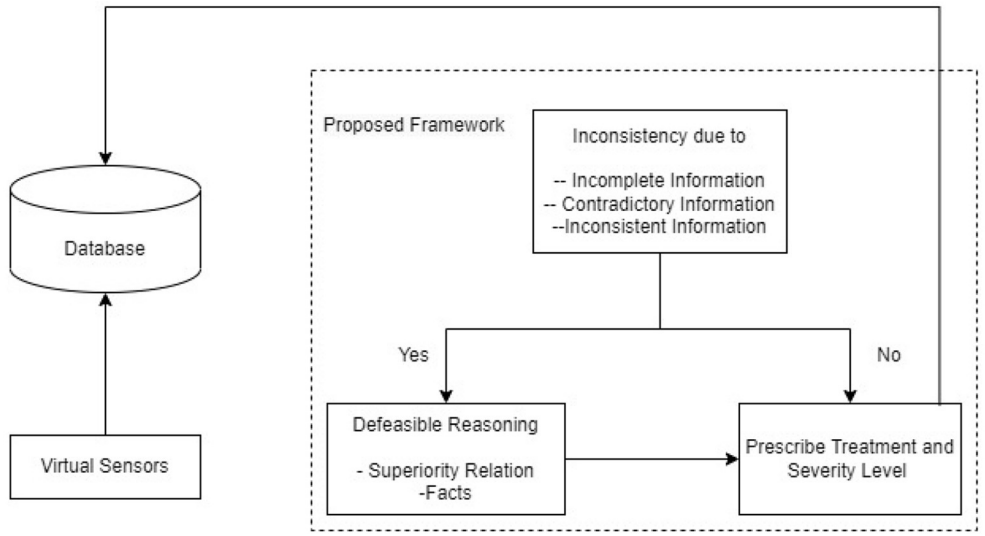


FIGURE 1 | Architecture diagram.

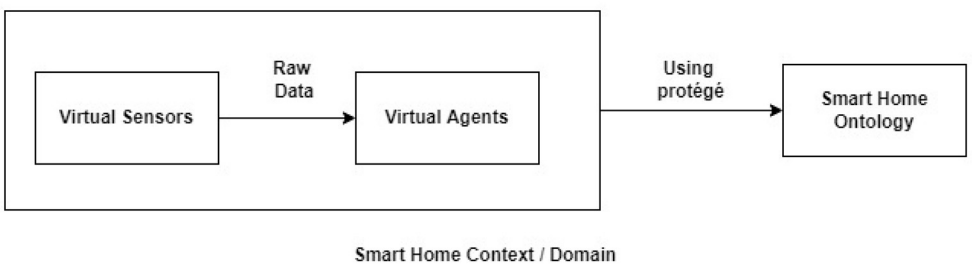


FIGURE 2 | Smart home system.

level and treatment, the system sends this information back to the database. The hospital has complete access to this database and will take action accordingly.

These systems, which are a smart home system and smart hospital system, then send their received data to the smart decision-making system embedded in the smart hospital system so that the smart decision-making system could form a diagnosis of the patient's current health and the necessary treatment. Since multiple heterogeneous contexts/systems/domains are involved in this framework, structuring the overall system and data is crucial. Therefore, ontology is used for this purpose. To conclude, the smart decision-making system needs some reasoning mechanism. For this purpose, contextual defeasible reasoning is used. The flow/sharing of information between heterogeneous contexts is done and maintained using rules. **Figure 2** explains the data acquisition and sharing in a smart home environment. The virtual sensors embedded in the smart home or worn by the patient inside the smart home collect raw data and send it to the virtual agents in a continuous manner. The virtual agents, which are intelligent, use their reasoning capabilities to generate an alert if a certain reading is obtained which is above or below a certain threshold. The agents alert the smart hospital domain's smart decision-making system.

Figure 3 explains the collection and transfer of data in a smart hospital environment. The virtual sensors embedded all over the hospital collect raw data and continuously share it with the virtual agents. The virtual agents use their reasoning capabilities and generate an alert if a particular acquired data reading is beyond or below a certain threshold. This alert generated is sent to the smart decision-making system by the agents implemented in the smart hospital domain. Databases containing information about the patient's allergies, and insurance, are also involved in this context or domain. They send the relevant data to the smart decision-making system whenever needed.

Intelligent virtual agents carry out the step that includes data acquisition of the MCS. The virtual agents take into account all data that real physical agents measure. The proposed system includes Multi-Context Systems consisting of several heterogeneous contexts. After this, in order to map the acquired data into the different domains of the system, ontologies are developed with their own set of rules. The developed ontologies are then mapped to their correlated description logic. Ontologies

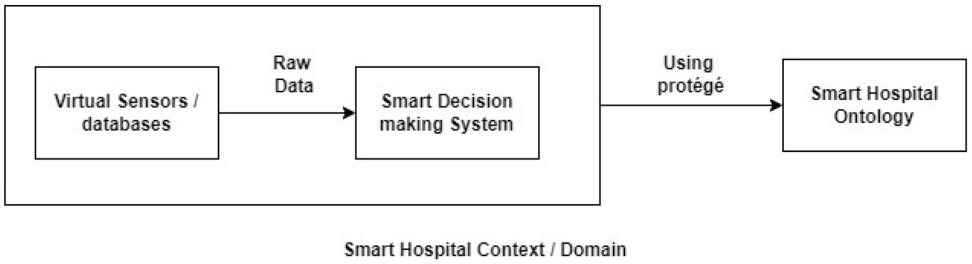


FIGURE 3 | Smart hospital system.

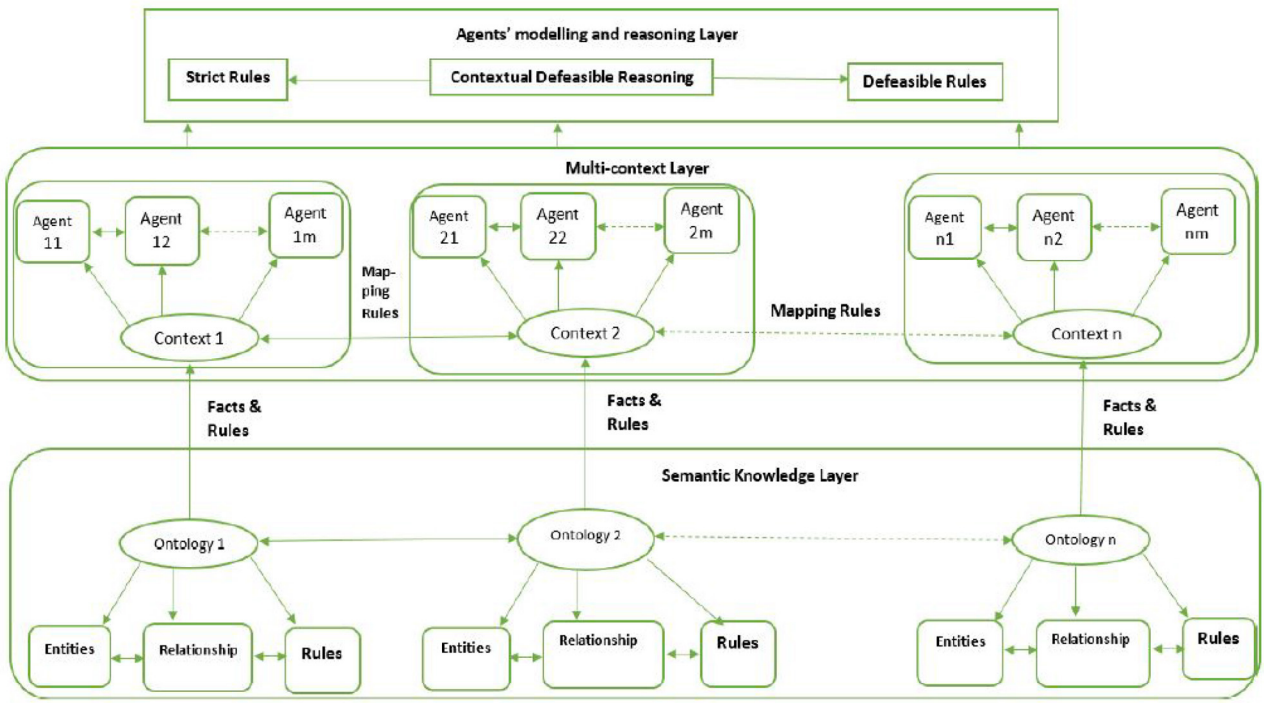


FIGURE 4 | Proposed system layered architecture.

based on descriptive logic use the distributed description logic for interconnection *via* mapping rules. Once the system gets settled, the flow of information begins by heterogeneous contexts interactivity. Contextual defeasible reasoning (CDL) is applied to the system after the information flow and used to handle any inconsistencies.

After this, contextual defeasible reasoning is applied to the system to handle any inconsistencies. CDL uses its facts, strict, defeasible, and defeater rules and preferences to solve these issues. The proposed framework verification is achieved by intending to develop a prototype or with the help of simulation. At the moment, implementing the proposed system is difficult due to financial and technical constraints. Our emphasis is mainly on research and partially on development. The development done is mainly for testing purposes. We are assuming that sensors are collecting the data. This data is modeled using agents. The reasoning performed on this system is done through contextual defeasible reasoning. Since our system is heterogeneous, therefore, initially, the agents are defined in different heterogeneous contexts. These heterogeneous contexts communicate with each other *via* mapping rules. **Figure 4** explains the overall layered architecture of the proposed framework.

4. FRAMEWORK DEVELOPMENT

4.1. Context Development

To check the correctness of this work, we have developed a case study about a person who has Parkinson's disease (a

Agent 11: Room Temperature Agent
Initial Facts: Person (‘Alice), Patient_ID(‘E001), has_Patient_ID(‘Alice, ‘E001), RoomTemp(’21), Location(‘Kitchen, ‘Room, ‘LivingRoom), hasRoomTemp (‘Location, ‘RoomTemp), HighRoomTemp(‘21, ‘?),LowRoomTemp(‘21, ‘?)
R111: person(?p), has_Patient_ID(?p, ?PID), Patient_ID(?PID) → Patient(?p) **R112:** Location(?L), hasRoomTemp (‘?L, ‘?rt), RoomTemp (‘?rt), HighRoomTemp(‘?rt, ‘35) →hasRTSituation (‘?RT, ‘high) **R113:** Location(?L), hasRoomTemp (‘?L, ‘?rt), RoomTemp (‘?rt), LowRoomTemp(‘?rt, ‘15) →hasRTSituation (‘?RT, ‘low)

FIGURE 5 | Example rules of context 1 smart home.

neurodegenerative disease). The following two contexts have been developed for this case study.

4.1.1. Context 1: Smart Home

Since memory loss is the dilemma of Parkinson's disease, Alice's house has been embedded with various sensors. One of them is the motion detector sensor for sensing the current position of Alice, i.e., whether she is sitting, standing, has fallen over, or is showing symptoms of Slowed movement (bradykinesia) (21) one of the cardinal manifestations of Parkinson's disease, and for determining the current location of Alice in her house, i.e., Alice is in which room of her house is currently. This sensor is embedded in every room of her house. There are also smoke detection sensors to detect the presence of smoke or fire in the house. Another sensor being used is the carbon monoxide (CO) detection sensor that helps in alerting about emergencies, such as a person has turned on the gas of the stove but forgot to light it. Carbon monoxide is an extremely gas, and interaction with it for an extended period can be injurious to health (22). The smoke detectors are embedded everywhere in the house. However, the carbon monoxide detector is embedded in the kitchen only. On average, the carbon monoxide level recorded in houses with gas stoves is between 0.5 and 5 parts per million (ppm). However, in houses with gas stoves that are properly fixed, the normal recorded level is between 5 and 15 ppm, but the level may be 30 ppm or higher if the gas stoves are not correctly fixed (22). **Figure 5** shows some example rules for the smart home context.

4.1.2. Context 1: Smart Hospital

In addition to home embedded sensors, some sensors are worn by Alice. These sensors are the heart rate monitoring sensor to detect fluctuations in heart rate, body temperature sensors for measuring changes in body temperature, and surface electrodes sensors for measuring electrocardiography (ECG), electromyography (EMG), or electroencephalography (EEG). They are used to monitor the electrical activities of the heart, muscles, and brain, respectively, directly from the skin's surface, eventually providing information about rigidity in the skin and detecting body tremors, both of which are prevalent symptoms of Parkinson's disease.

It must be mentioned here that a specific range of "normal" values is associated with each of these sensors. In case a reading value of a sensor goes above or below this range, an alert is issued by the sensor. The normal range for a heart monitoring sensor is 60–100 bpm (precisely 82 bpm). Usually, an adult has a body temperature between 97 and 99 F (22). For ECG sensors, normal electrocardiography intervals are 0.6–1.2 s. Based on the muscle, the normal EMG range is between 50 μV and 30 mV. An average adult has EEG readings between 8 Hz and higher. For EMG, a reading of 7 Hz or less is considered abnormal in awake adults. However, they are considered normal in children, or sleeping adults (23). **Figure 6** shows some example rules for the wearable device context.

4.2. Agent Development

Nine intelligent virtual agents are designed for the proposed framework. Their description is provided below.

- Systolic Blood Pressure (BP) Agent: For monitoring the systolic BP
- Diastolic BP Agent: For monitoring the diastolic BP
- Body Temperature Agent: For monitoring the body temperature
- Heart-Rate-Agent: For monitoring the Heart Rate
- EEG-Agent: For monitoring the EEG
- ECG-Agent: For monitoring the ECG
- EMG-Agent: For monitoring the EMG
- Carbon Monoxide-Agent: This agent monitors the carbon monoxide level
- Carbon Dioxide-Agent: This agent monitors the carbon dioxide level
- Room-Temperature-Agent: This agent monitors the room temperature

4.3. Ontology Development

For system modeling, contextual information is obtained from multiple ontologies.To model distributed domains for the

Agent 21: Body Temperature Sensor
Initial Facts: Person (‘Alice), Patient_ID(‘P002), has_Patient_ID(‘Alice, ‘P002), Body_Temperature(‘99.5), has_Body_Temperature(‘Alice, ‘99.5), Greater_Than_Equal(‘99.5, ‘100.9), Less_Than_Equal(‘99.5, ‘95.0)
R211: person(?p), has_Patient_ID(?p, ?PID), Patient_ID(?PID) →Patient(?p) **R212:** person(?p), has_Body_Temperature(‘?p, ‘?temp), Body_Temperature(‘?temp), Greater_Than_Equal(‘?temp, ‘100.9) → has_Fever(‘?p, ‘high) **R213:**has_Fever(‘?p, ‘high) → Tell (31, 54, (has_Fever(‘?p, ‘high)) **R214:** person(?p), has_Body_Temperature(‘?p, ‘?temp), Body_Temperature(‘?temp), Less_Than_Equal(‘?temp, ‘95.0) → has_Hypothermia(‘?p, ‘high) **R215:**has_Hypothermia(‘?p, ‘high) → Tell (21, 54, (has_Hypothermia(‘?p, ‘high)) **R216:** person(?p), has_Body_Temperature(‘?p, ‘?temp), Body_Temperature(‘?temp), ~Greater_Than_Equal(‘?temp, ‘100.9), ~Less_Than_Equal(‘?temp, ‘95.0) →Current_Body_Temperature(‘?p, ‘normal)

FIGURE 6 | Example rules of context 2 smart hospital.

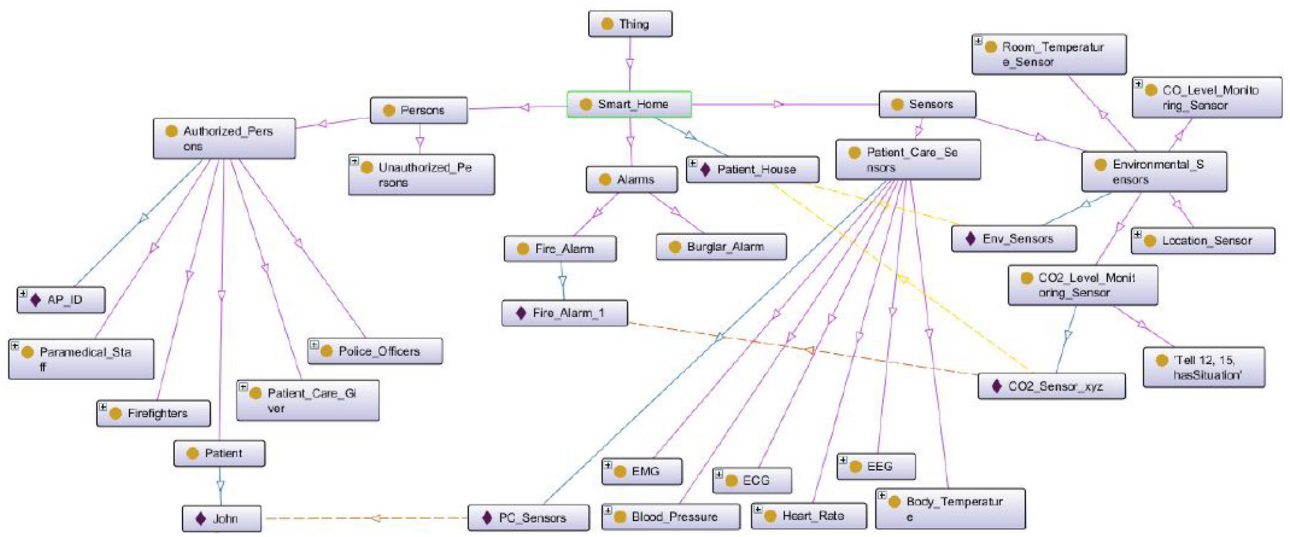


FIGURE 7 | Smart home ontology.

proposed system, a smart home ontology and a smart hospital ontology are developed for modeling the proposed framework. Fragment of the smart home ontology is shown in **Figure 7**.

The smart home ontology describes the flow of information in this context. The second ontology developed is the smart hospital ontology. This ontology describes the structuring of the smart hospital context and expresses the relations between different smart hospital context entities. It also describes the flow of information in this context. Fragment of the smart hospital ontology is shown in **Figure 8**.

In the proposed framework, agents acquire contextual data either from a single ontology or multiple ontologies based on the system's design. Mapping rules have been designed to model the flow of information between the contexts involved in the framework. To manage inconsistencies in such systems, priorities are assigned to the mapping rules, with the rule having the

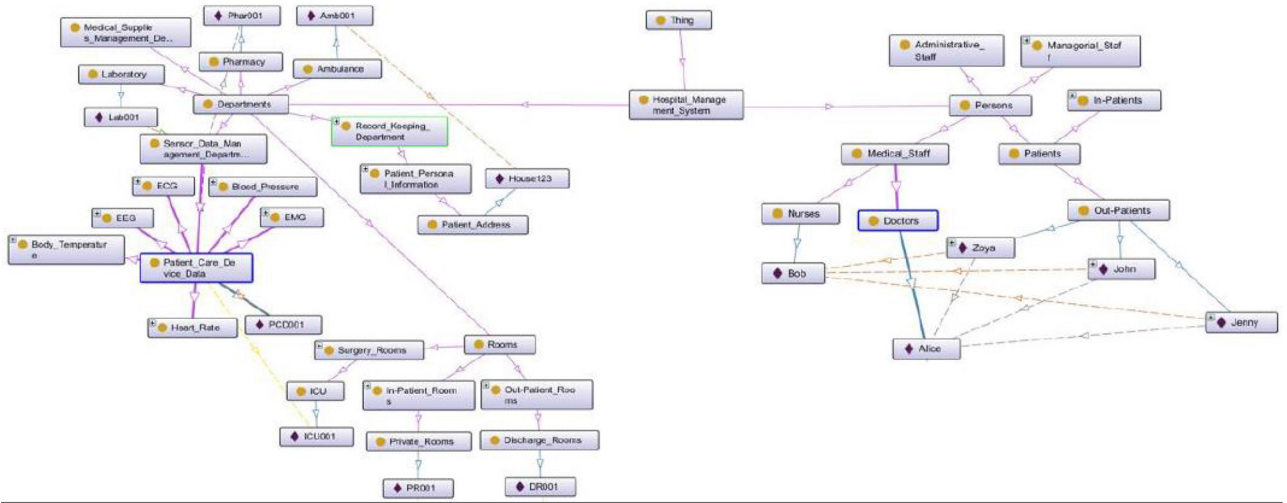


FIGURE 8 | Smart hospital ontology.

higher priority being fired. Class hierarchies of the developed two ontologies are also shown in **Figure 7**.

5. TEMPORAL LOGIC FORMALISM

There are two types of models in temporal logic. Linear temporal logic (LTL) and computational tree temporal logic (CTL*) (24). In our work, we use CLT* instead of LTL because LTL is based on a single computational path (25).

The semantic of TL is defined by CLT* (26). The changing of the current state, i.e., state to -> state from and state from -> state to, is represented by the ω-tree shape (27). The states correspond to the agent knowledge base and communication mechanism. The agent fires multiple rules in a multi-agent system to reach the required state that is acceptable as an outcome. Each agent may drive new context whenever it matches the pre-defined rules. When the agent acts on its actions, the system moves to the next state. If the agent did not acquire the data, the agent tells its corresponding agent about the context, and then the system moves to the other state.

5.1. Current Situation Observation

When an agent gets something from its corresponding sensor, it checks it to map the acquired context with the internal knowledge.

5.1.1. Pre-defined Rules

- G Ag^i (ϕ, t1) where $\phi \in IK^i$ - (i)
- G Ag^i ($\neg\phi$, t1) where $\phi \in IK^i$ - (ii)

In the rule mentioned above i and ii, globally, agent i gets information from the environment which belongs to the agent's internal knowledge IK or information has not been detected by agent i but belongs to internal knowledge. Here, IK^i represents internal knowledge. For instance, when the fire information is acquired to the agent thus the fire context belongs to the agent's internal knowledge; else, not fire belongs to it.

- F [$\ni$ Tell (i, j, ϕ)] - (iii)

In rule iii, it states in the future, on some existing states s' $agent^i$ believes in telling $agent^i$ about some acquired context information, so the $agent^i$ can be auto-trigger.

- Ag^i (ϕ, t1) where $\phi \in$ HI - (iv)

$Agent^i$ acquired the state formula concerning t1, and it is true, and that state formula belongs to acquire HI. For example, after selecting the optimal plan, $agent^i$ has the information about calling the humanitarian assistance for human involvement to give the alert related to the unauthorized entity, as shown in rule iv.

5.1.2. Successful Events

- ϕ i ε ψ i - (v)

When $agent^i$ acquires the context ϕ about one state formula or $agent^i$ acquires the context ψ about other state formula or both. In every way, it would be a successful case. For example, if $agent^i$ gets the information about an animal who is on-site where the camp of soldiers are, or $agent^i$ gets the information about detecting the unauthorized entity or both. It would be a successful case, as shown in rule v.

- Agi ($\neg\phi$, t1) where $\phi \epsilon IK^i$ -(vi)

Agent i has not detected any information, and it does belong to internal knowledge; therefore, it is the case of partial success, e.g., an agent did not get the information about detecting any unauthorized entity, is still a partial success case, as shown in rule vi.

- Ask (j, i, ϕ) ε Tell (i, j, ϕ) - (vii)

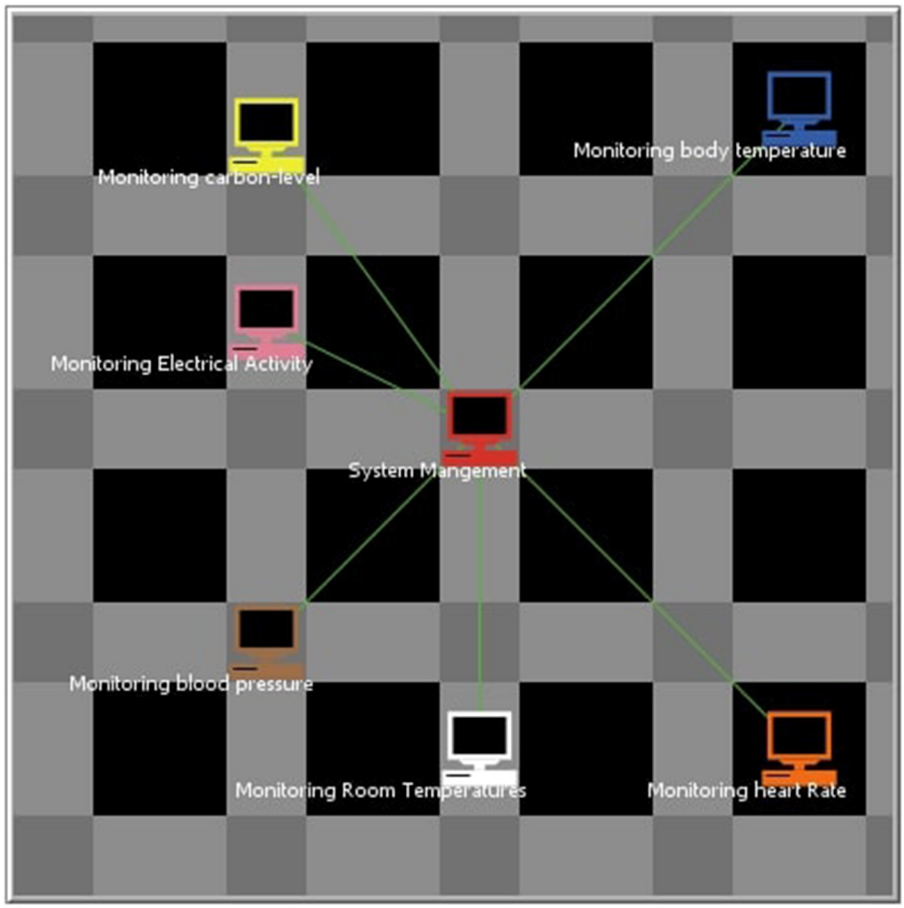


FIGURE 9 | System interface.

In rule vii, agent^i can ask agent^i about the specific context, or agent^i has the pre-defined rule to tell the specific context to agent^i. For example, agent^i has the rule to ask agent^i about detection, or agent^i has the rule to tell agent^i about the detected entity.

5.1.3. Current Situation Step of Actions

After the agent checks the acquired information is true or not, it has the step of actions to follow to act efficiently.

- Xi (α, ti) - (viii)

In rule viii, after matching the predefined rules, the very next thing is, agent^i has to perform θ certain actions α. For instance, if the unauthorized entity context detection is true, then agent^i has actions α to perform θ in a specific time t_i.

- F [$\ni$ i (α (HI), t_i)] - (ix)

After matching the pre-defined rules, in the future, in some existing states, an agent^i has the actions to perform HI requirements. After selecting the optimal plan, agent^i would contact the HI to prevent the situation from any damage, as shown in rule xi.

5.1.4. Task Priorities

- DL (H) iff ϕ i $\cup$ ψ j - (x)
- DL (S) iff ϕ i $\wedge$ ψ j - (xi)

The danger level DL is always high (H). If the one-state formula is detected ϕ unless the other state formula ε is detected, DL changes to severe (S). For example, unauthorized entity ϕ with any weapon at the site ψ is severe DL, as shown in rules x and xi.

6. SIMULATION OF PROPOSED SYSTEM

This section explains the case study, we have developed for the validation of our system. For its implementation, we have used the Netlogo simulation tool. Netlogo is used for modeling, and it teaches the concept of agents like turtles, patches, observers and links, etc., it is used in different kinds of scenarios like in gaming, or if people built a model for a war to check it whether it is correct or not, or/ and in disaster management scenarios, etc. It is open-source software that includes an interface, commands to create an agent model and execute them. It has built-in commands and reports that they call primitive, and the command reports created

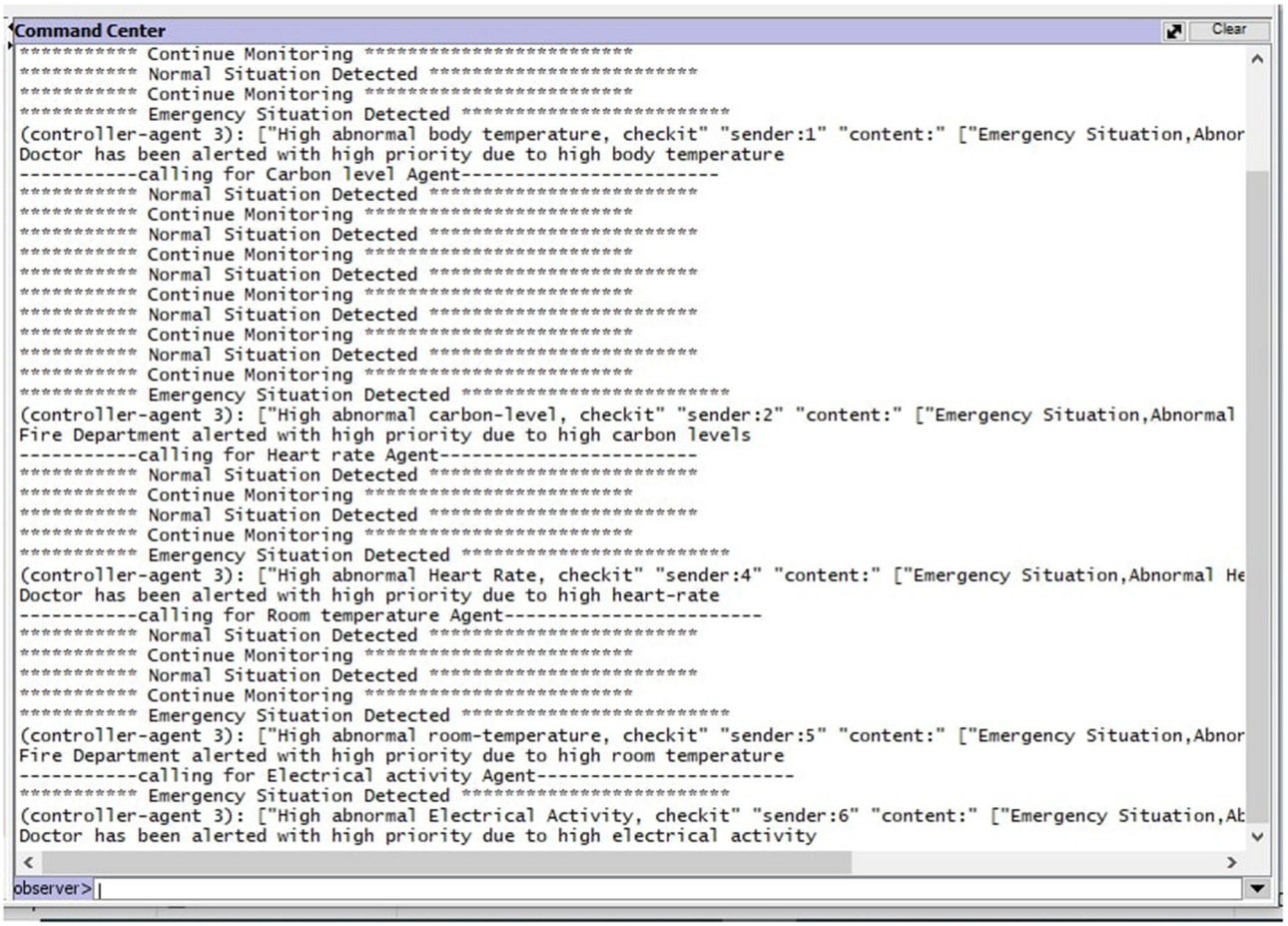


FIGURE 10 | System interface.

by a programmer called procedure (28, 29). To achieve the perfect communication between agents, FIPA protocol was used (30).

For example, agent 1 tells agent two about an "Emergency Situation" by sending "Abnormal BP detected" in the message.

["Emergency Situation" "sender:1""content:" "Abnormal BP Detected" "receiver:2."]

Figure 9 shows the interface of our system. We created seven context-aware agents named controller agent, BP agent, body temperature agent, heart-rate agent, electrical-activity agent, carbon-level agent, room-temperature agent. Each of them has their unique functionality based on their own set of facts and rules, also perform reasoning and take decisions intelligently. These agents can communicate in NetLogo as they are connected by creating a link.

Figure 10 shows the execution of the system. During the execution, the system continuously monitors the patient's situation as "Continuously Monitoring." When any "Emergency Situation" is detected controller agent gets to inform a respective agent that is highly abnormal body temperature detected. After that controller agent sends a message to the doctor by alerting the situation "Doctor has been alerted with high priority due to high body temperature."

7. ALGORITHM

The algorithm has two parts. The first part (**Algorithm 1**) is invoked as soon as the system begins. However, invoking the second part of the algorithm (**Algorithm 2**) depends on obtaining an abnormal value (above or below the pre-defined normal range). In **Algorithm 1**, the agents collect contextual information from their environment. Until that contextual information remains within a specific range (the value obtained is above the lower pre-defined range and below the upper pre-defined range), the system generates no alert and usually works. However, if any agent obtains a value that does not fall within the pre-defined normal range (the value obtained is either below the lower or above the upper pre-defined range), the system generates an alert (**Algorithm 2**) invoked.

In **Algorithm 2**, the priority of the rule instances is fired, and the rule with the highest priority is chosen. The following action, alerting the corresponding authority, depends on the rule instance fired. In the case of PC_Agents rule instances, the

Algorithm 1: Obtaining of Data.

Input:

- Contextual_Info^{xi}: contextual data acquired by an agent^{i} from the environment
- high_value^{i}: the maximum value of given range
- low_val^{i}: the minimum value of given range
- flagVar: flag will be 0 when system continues and 1 and when the system stops
- r^{i}: the fired rule instance

Output: Alert when out of the given range Contextual_info$^{i}_{x}$ is obtained
Start ∀ *Agent*^{i} ∈ agents{*n*}
∃ (*Env_Agents*^{i} ∈ agentsn || PC_Agents^{i} ∈ *agents*{*n*})
if (Contextual_Info^{xi} < high_val^{i} && Contextual_Info^{xi} > low_val^{i}) **then**
  return Contextual_Info^{xi} is normal
else
  invoke Algorithm 2 (Situation abnormal)
end
if some act done **then**
  flagVar = = 1
else
  flagVar = = 0
end
if Flag = 1 **then**
  System Stops
else
  Flag = 0
  System continues as normal
  Invoke Algorithm 1
end

doctors, caretakers, and the next of kin are alerted. On the other hand, when the rule instances fired are from Env_Agents, then the fire department, emergency response unit, and the police department are alerted.

This alert is sent to the corresponding authority by the controller agent of the system. The controller agent is the primary agent of the system controlling all information flow. All the contextual information obtained by any agent is sent directly to the controller agent. In case of an emergency (an abnormal value), the agent sends an alert to the controller agent, and the controller agent then contacts the corresponding authority. The contact can be made in any number of ways, and it can be a text message, an email, a notification, a phone call, or a WhatsApp message. The mode of communication is chosen by the corresponding authority and then included as a command in the memory of the controller agent.

Once an alert has been generated, it is up to the corresponding authority to either take action or not do anything. The corresponding authority solely takes this decision. The reason for this decision goes beyond the scope of this work and the author's domain. If the corresponding authority believes action must be taken, it alerts the system, and the flag's value is set to 1. The flag is a variable that can have only two values, either 1 or 0. each value corresponds to a separate course of actions. If the flag's value becomes 1, the system knows that the corresponding authorities have taken over. In such a case, the system will halt and wait to be rebooted by the corresponding authority. However, if the corresponding authority believes that no action needs to be taken, it does not alert the system, and the flag's value is set to 0. If the flag's value becomes 0, then the system knows that the corresponding authorities have not taken over now, and they will continue to work as usual (continue gathering contextual information). In such a case, the system will not halt. The working of the algorithm in case a single rule is fired or in case multiple rules are fired simultaneously is described below. It must be mentioned here that by default, the priority of patient care agents (PC_Agents) is more than environmental agents (Env_Agents).

7.1. Case 1: Single Rule Instance Fired

When a single rule is fired, the system will function according to that rule without any conflict or issue.

7.2. Case 2: Multiple Rules Instances

When multiple rules are fired, then the set CRS_{pri} will be developed. All the conflicting rules will be included in the CRS_{pri} set, and ultimately the rule with the highest priority will be selected.

8. LIMITATIONS AND FUTURE SCOPE

Since the inconsistencies in heterogeneous knowledge sources systems usually do not have a single reason for occurrence but arrive from the interaction of heterogeneous knowledge bases, therefore, conflicting, or contradictory information is a major

Algorithm 2: Conflicting Rule Set Creation

Input:

- Contextual_Info^{xi}: contextual data acquired by an agent^{i} from the environment
- high_value^{i}: the maximum value of the given range
- low_val^{i}: the minimum value of the given range
- flagVar: flag will be 0 when system continues and 1 and when the system stops
- r^{i}: the fired rule instance
- r$_{pri}{}^{i}$: the priority of the rule
- r^{j}: the other fired rule instance
- r$_{pri}{}^{j}$: the priority of the other rule
- CRS^{pri}: Prioritized rules set of conflicting rules
- CRS^{ord}: Ordinary rules set of conflicting rules

Output: CRS with all those rules conflict the fired rule

Start
if r^{xi} && r^{j} ∈ PC_Agents **then**
  r$_{pri}{}^{i}$ == r$_{pri}{}^{j}$ == prioritized CRS^{pri} = m ← r^{i} && r^{j} **if** r$_{pri}{}^{i}$ > = r$_{pri}{}^{j}$ **then**
    Return r^{i}
  else
    Return r^{j}
  end
else
  invoke Algorithm 2 (Situation abnormal)
end
if r^{xi} && r^{j} ∈ Env_Agents **then**
  r$_{pri}{}^{i}$ == r$_{pri}{}^{j}$ == ordinary CRS^{ord} = m ← r^{i} && r^{j} **if** r$_{ord}{}^{i}$ > = r$_{ord}{}^{j}$ **then**
    Return r^{i}
  else
    Return r^{j}
  end
else
  invoke Algorithm 2 (Situation abnormal)
end
if r^{xi} ∈ PC_Agents && r^{j} ∈ Env_Agents **then**
  r$_{pri}{}^{i}$ == prioritized
  CRS^{pri} = m ← r^{i}
  r$_{pri}{}^{j}$ == ordinary
  CRS^{ord} = m ← r^{j}
  Return r^{i}
else
  invoke Algorithm 2 (In case of a situation that is not normal)
end
if some act done **then**
  flagVar = = 1
else
  flagVar = = 0
end

reason for an inconsistent system. According to the best of our knowledge, the existing approaches propose no hard and fast method for assessing inconsistencies in MCS as so far, no proposed method is flexible enough to accommodate the criteria of various applications. A new approach is needed to provide a single platform to overcome the issues of inconsistency due to conflicting or contradictory information. This new approach must have a reasoning mechanism to handle the issues while having a solid logical base to perform sound reasoning. Since these issues occur due to knowledge sharing among heterogeneous contexts, a mechanism is needed to represent the knowledge formally. There can be multiple heterogeneous

contexts in a multi-context system. Thus having a mechanism for structuring such a system in an organized manner is extremely necessary. The proposed framework fulfills all the requirements mentioned above by using ontology to structure the system, description logic for knowledge representation, and contextual defeasible reasoning for performing reasoning. This work is beneficial for patients who require continuous monitoring of their health, but they cannot afford to stay in a monitoring facility or do not have enough resources to accommodate all their patients. In such cases, this system can be helpful as it will allow the patient to be under supervision continuously by the agents, and in case of an emergency, the authorities will be instantly notified. Another implementation of this system is for those patients who may not need continuous monitoring but are suffering from terminal illnesses such as Parkinson's disease. In these cases, the disease can act up at any moment. The exact time of such situations cannot be predicted. In these scenarios, the patients can be monitored in the safety of their homes, and in case of emergencies, the medical authorities will be notified immediately. In other domains, this system can also be implemented. In disaster management systems, this system can be used to manage disasters such as fire eruptions, carbon dioxide or carbon monoxide leakage, the occurrence of flood, etc.

9. CONCLUSION

In this work, an ontology-driven formalism has been proposed for handling inconsistency in a highly dynamic environment. The proposed framework and application differ from other approaches because it uses contextual defeasible logic to solve inconsistencies. The proposed framework can provide a single platform to overcome the issues of inconsistency occurring due to contradictory information in a multi-context system. Contextual defeasible reasoning has been chosen to perform reasoning on the system as it can handle inconsistency issues by providing a solid logical base to perform sound reasoning. Since the issue occurs due to sharing knowledge among heterogeneous contexts, description logic (DL) and distributed description logic (DDL) mechanisms represent the knowledge being shared formally. While an ontology-based modeling approach has been adopted to model the multiple heterogeneous contexts in a multi-context system. Thus, the proposed framework provides a formalism of multi-agents based on that can handle conflicting information in an environment that is highly decentralized in nature.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AJ and SA: conceptualization. SA: data curation. AJ: formal analysis, investigation, methodology, and software. SB and MN: funding acquisition. KS and MN: project administration. RA and MN: resources and writing—review and editing. AM, AJ, and KS: supervision. SA, AM, and RA: validation. SA and KS: visualization. All authors contributed to the article and approved the submitted version.

FUNDING

This paper was supported by TU-Dresden, Dresden, Germany.

REFERENCES

1. Akhtar SM, Nazir M, Saleem K, Haque HMU, Hussain I. An ontology-driven IoT based healthcare formalism. *Int J Adv Comput Sci Appl.* (2020) 11:479–86. doi: 10.14569/IJACSA.2020.0110261
2. Akhtar SM, Haque HMU. Contextual defeasible reasoning framework for heterogeneous systems. In: *Context-Aware Systems and Applications, and Nature of Computation and Communication.* Springer (2020). p. 16–30.
3. Javed AR, Sarwar MU, ur Rehman S, Khan HU, Al-Otaibi YD, Alnumay WS. Pp-spa: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Process Lett.* (2021) 10:1–18. doi: 10.1007/s11063-020-10414-5
4. Mahfooz ul Haque H, Muhammad Akhtar S, Uddin I. Contextual defeasible reasoning framework for heterogeneous knowledge sources. In: *Concurrency and Computation: Practice and Experience*. London: Springer-Verlag. p. e6446.
5. Sharmeen Z, Martinez-Enriquez AM, Aslam M, Syed AZ, Waheed T. Multi agent system based interface for natural disaster. In: *International Conference on Active Media Technology*. Warsaw: Springer (2014). p. 299–310.
6. Haiouni H, Maamri R. Context-aware multi-agent planning in intelligent environments. *Multiagent Grid Syst.* (2019) 15:219–36. doi: 10.3233/MGS-190310
7. Esposito A, Tarricone L, Zappatore M, Catarinucci L, Colella R. A framework for context-aware home-health monitoring. *Int J Auton Adapt Commun Syst.* (2010) 3:75–91. doi: 10.1007/978-3-540-69293-5_11
8. Al-Anbaki NS, Obeid N, Sabri KE. A defeasible logic-based framework for contextualizing deployed applications. *Work.* (2019) 10:176–86. doi: 10.14569/IJACSA.2019.0100923
9. Eiter T, Fink M, Schüller P, Weinzierl A. Towards diagnosing inconsistency in nonmonotonic multi-context systems. *Logic Based Interpret.* (2009) 9:13. doi: 10.1007/978-3-642-15675-5_14
10. Bellini P, Benigni M, Billero R, Nesi P, Rauch N. Km4City ontology building vs data harvesting and cleaning for smart-city services. *J Vis Lang Comput.* (2014) 25:827–39. doi: 10.1016/j.jvlc.2014.10.023
11. Tapakire BA, Patil MM. IoT based smart agriculture using thingspeak. *Int J Eng Res Technol.* (2019) 8. doi: 10.17577/IJERTV8IS120185
12. Maddikunta PKR, Pham QV, Prabadevi B, Deepa N, Dev K, Gadekallu TR, et al. Industry 5.0: a survey on enabling technologies and potential applications. *J Ind Inf Integr.* (2021) 15:100257. doi: 10.1016/j.jii.2021.100257
13. Shrivastava R, Kumar P, Tripathi S, Tiwari V, Rajput DS, Gadekallu TR, et al. A novel grid and place neuron's computational modeling to learn spatial semantics of an environment. *Appl Sci.* (2020) 10:5147. doi: 10.3390/app10155147
14. Bhattacharya S, Maddikunta PKR, Hakak S, Khan WZ, Bashir AK, Jolfaei A, et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimedia Tools Appl.* (2020) 10:664–70. doi: 10.1007/s11042-020-09988-y
15. Wang W, Xu H, Alazab M, Gadekallu TR, Han Z, Su C. Blockchain-based reliable and efficient certificateless signature for IIoT devices. *IEEE Trans Ind Inf.* (2021) 2:193–200. doi: 10.1109/TII.2021.3084753

16. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet Things J.* (2021) 1–18. doi: 10.1109/JIOT.2021.3117762
17. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inf.* (2021) 42:4–13. doi: 10.1109/JBHI.2021.3112693
18. Haque HMU, Khan SU, Hussain I. Semantic knowledge transformation for context-aware heterogeneous formalisms. *Int J Adv Comput Sci Appl (IJACSA).* (2019) 10:664–70. doi: 10.14569/IJACSA.2019.0101285
19. Eiter T, Fink M, Weinzierl A. Preference-based inconsistency assessment in multi-context systems. In: *European Workshop on Logics in Artificial Intelligence.* Helsinki: Springer. (2010). p. 143–55.
20. Haque HMU, Khan SU. A context-aware reasoning framework for heterogeneous systems. In: *2018 International Conference on Advancements in Computational Sciences (ICACS).* Lahore: IEEE (2018). p. 1–9.
21. Johnson ME, Stecher B, Labrie V, Brundin L, Brundin P. Triggers, facilitators, and aggravators: redefining Parkinson's disease pathogenesis. *Trends Neurosci.* (2019) 42:4–13. doi: 10.1016/j.tins.2018.09.007
22. Mshali H, Lemlouma T, Moloney M, Magoni D. A survey on health monitoring systems for health smart homes. *Int J Ind Ergon.* (2018) 66:26–56. doi: 10.1016/j.ergon.2018.02.002
23. Yassine A, Singh S, Alamri A. Mining human activity patterns from smart home big data for health care applications. *IEEE Access.* (2017) 5:13131–41. doi: 10.1109/ACCESS.2017.2719921
24. Alagar V, Periyasamy K. Temporal logic. In: *Specification of Software Systems.* Sangli: Springer (2011). p. 177–229.
25. Goranko V, Rumberg A. Temporal logic. In: *Stanford Encyclopedia of Philosophy.* Evanston: Center for Connected Learning and Computer-Based Modeling Northwestern University (2020).
26. van der Aalst WM, De Beer H, van Dongen BF. Process mining and verification of properties: an approach based on temporal logic. In: *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems".* Eindhoven: Springer (2005). p. 130–47.
27. Dormoy J, Kouchnarenko O, Lanoix A. Using temporal logic for dynamic reconfigurations of components. In: *International Workshop on Formal Aspects of Component Software.* Berlin: Springer. (2010). p. 200–17.
28. Tisue S, Wilensky U. NetLogo: design and implementation of a multi-agent modeling environment. In: *Proceedings of Agent. Vol. 2004* (Fukushima). (2004). p. 7–9.
29. Sakellariou I, Kefalas P, Stamatopoulou I. Enhancing NetLogo to simulate BDI communicating agents. In: *Hellenic Conference on Artificial Intelligence.* Berlin: Springer (2008). p. 263–75.
30. Hong Mw, Cho Dj. Ontology context model for context-aware learning service in ubiquitous learning environments. *Int J Comput.* (2008) 2:172–8. doi: 10.1145/2590651.2590776

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the author AJ.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Akhtar, Nazir, Saleem, Ahmad, Javed, S. Band and Mosavi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2022.819156

COVID-19 Identification System Using Transfer Learning Technique With Mobile-NetV2 and Chest X-Ray Images

Mahmoud Ragab[1,2,3]*, *Samah Alshehri*[4], *Gamil Abdel Azim*[5], *Hibah M. Aldawsari*[6,7], *Adeeb Noor*[1], *Jaber Alyami*[8,9] and *S. Abdel-khalek*[10,11]

[1] Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, [2] Centre of Artificial Intelligence for Precision Medicines, King Abdulaziz University, Jeddah, Saudi Arabia, [3] Mathematics Department, Faculty of Science, Al-Azhar University, Naser City, Egypt, [4] Department of Pharmacy Practice, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, [5] Department of Computer Science, Faculty of Computers and Information Canal Suez University, Ismalila, Egypt, [6] Department of Pharmaceutics, Faculty of Pharmacy, King Abdulaziz University, Jeddah, Saudi Arabia, [7] Center of Excellence for Drug Research and Pharmaceutical Industries, King Abdulaziz University, Jeddah, Saudi Arabia, [8] Department of Diagnostic Radiology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, [9] Imaging Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia, [10] Department of Mathematics, Faculty of Science, Taif University, Taif, Saudi Arabia, [11] Department of Mathematics, Faculty of Science, Sohag University, Sohag, Egypt

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Abdul Rehman Javed, Air University, Pakistan
Praveen Kumar, VIT University, India

***Correspondence:**
Mahmoud Ragab
mragab@kau.edu.sa

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *20 November 2021*
Accepted: *12 January 2022*
Published: *03 March 2022*

Citation:
Ragab M, Alshehri S, Azim GA, Aldawsari HM, Noor A, Alyami J and Abdel-khalek S (2022) COVID-19 Identification System Using Transfer Learning Technique With Mobile-NetV2 and Chest X-Ray Images. Front. Public Health 10:819156. doi: 10.3389/fpubh.2022.819156

Diagnosis is a crucial precautionary step in research studies of the coronavirus disease, which shows indications similar to those of various pneumonia types. The COVID-19 pandemic has caused a significant outbreak in more than 150 nations and has significantly affected the wellness and lives of many individuals globally. Particularly, discovering the patients infected with COVID-19 early and providing them with treatment is an important way of fighting the pandemic. Radiography and radiology could be the fastest techniques for recognizing infected individuals. Artificial intelligence strategies have the potential to overcome this difficulty. Particularly, transfer learning MobileNetV2 is a convolutional neural network architecture that can perform well on mobile devices. In this study, we used MobileNetV2 with transfer learning and augmentation data techniques as a classifier to recognize the coronavirus disease. Two datasets were used: the first consisted of 309 chest X-ray images (102 with COVID-19 and 207 were normal), and the second consisted of 516 chest X-ray images (102 with COVID-19 and 414 were normal). We assessed the model based on its sensitivity rate, specificity rate, confusion matrix, and F1-measure. Additionally, we present a receiver operating characteristic curve. The numerical simulation reveals that the model accuracy is 95.8% and 100% at dropouts of 0.3 and 0.4, respectively. The model was implemented using Keras and Python programming.

Keywords: machine learning, convolution neural networks, transfer learning, MobileNetV2, COVID-19

INTRODUCTION

The coronavirus disease (COVID-19) has endangered social lives and created disturbing financial costs. Many studies have attempted to find a way to manage its spread and the resulting death. Moreover, many research propositions have been made to evaluate the existence and seriousness of pneumonia triggered by COVID-19 (1–4). Compared to CT, although radiography is easily accessible in hospitals worldwide, X-ray images are considered less delicate for examining patients with COVID-19. Primary analysis is crucial for instant seclusion of the infected individuals. Moreover, it decreases the rate of infection in a healthy population because of the accessibility of sufficient therapy or vaccination for the virus.

Chest radiography is an essential method for identifying pneumonia. This is easy to accomplish with rapid medical identification. Upper body CT has a high level of susceptibility for the identification of COVID-19 (5). The X-ray image reveals aesthetic keys associated with the virus (6, 7). They considered the tremendous rate of suspected individuals and a minimal variety of trained radiologists. These automated identification methods with refined abnormalities help diagnose diseases and raise early medical identification quality with high accuracy.

A therapeutic image in the form of a chest X-ray is crucial for the automatic clinical identification of COVID-19. A clinical approach for identifying COVID-19 based on artificial intelligence (AI) can be as precise as a human, conserve the radiotherapist period, and perform clinical identification less expensively and faster than standard techniques. AI explanations can be influential approaches for dealing with such problems. A machine learning structure was used to predict COVID-19 from a chest X-ray image. In contrast to classic medical image category approaches, which stick to a two-phase treatment (handmade function abstraction plus identification), we used an end-to-end deep determining structure that directly discovers COVID-19 from raw pictures without any pre-processing (8–14). Deep learning (DL)-based variations, specifically convolutional neural networks (CNNs), have been shown to perform better than timeless AI approaches. Recently, various processing system ideas and clinical image evaluation tasks have been utilized in several issues, including categorization, dissection, and face identification (15–19).

Convolutional neural network (CNN) is considered among the finest therapeutic imaging applications (20), particularly for categorization. CNN is better suited for large databases and requires computational assets (storage and memory). In most cases (as in this research), the database is inadequate; therefore, it is insufficient for building and training a CNN. Thus, obtaining the benefit of the CNN's power and transfer learning can reduce the computational cost (21, 22). A proficient CNN on a large and varied image database can perform specific classification tasks (23). Hence, many pre-trained designs such as VGG-Net, (24), Resnet (25), NAS-Net (26), Mobile-Net (27), Inceptions (28), and Xception (29) have won many of the world image classification competitions.

Recently, Bhattachary et al. in (30, 31) presented a survey of most of the DL models in the last five years, used to identify COVID-19 with different data sets. Kermany et al. (32) used the DL framework with transfer learning, Rajaraman et al. (33) used a customized CNN model- VGG16, Wang, et al. (34) proposed AI and DL based Frameworks, Shan et al. (35) presented Human-In-the-Loop Strategy and DL Based Segmentation Network VBNet, Ghoshal and Tucker (36) used Drop weights-based Bayesian Convolutional Network, Apostolopoulos and Mpesiana (37) presented Transfer Learning based on CNN, Esmail et al. (38) used CNN Architecture. Khan et al. (39) presented COVIDX-Net comprising Deep CNN Models. Sahiner et al. (40) proposed Deep CNN Models – ResNet50, InceptionV3, and Inception-ResNetV2.

How can the researchers in poor and developing countries quickly and appropriately contribute to stopping the spread of the virus and thus contributing to the economy's growth? Using AI and machine learning models with transfer learning

TABLE 1 | Mobile-NetV2 blockage.

Input	Operator	Output
W × H × N	1 × 1 2D convolution (conv2d) layer, ReLU6	W × H × tN
W × H × tN	3 × 3 dwise s = s ReLU6	W/s × H /s × tN
W/s × H/s × tN	Linear 1 × 1 2D convolution (conv2d) layer	W/s × H/s × M

TABLE 2 | Total network framework of Mobile-NetV2.

Input Shape	Operator	t	c	n	s
224 × 224 × 3	conv2d	-	32	1	2
112 × 112 × 32	bottleneck	1	16	1	1
112 × 112 × 16	bottleneck	6	24	2	2
56 × 56 × 24	bottleneck	6	32	3	2
28 × 28 × 32	bottleneck	6	64	4	2
14 × 14 × 64	bottleneck	6	96	3	1
14 × 14 × 96	bottleneck	6	160	3	2
7 × 7 × 160	bottleneck	6	320	1	1
7 × 7 × 320	conv2d 1 × 1	-	1280	1	1
7 × 7 × 1280	avgpool 7 × 7	-	-	1	-
1 × 1 × 1280	conv2d 1 × 1	-	k	-	

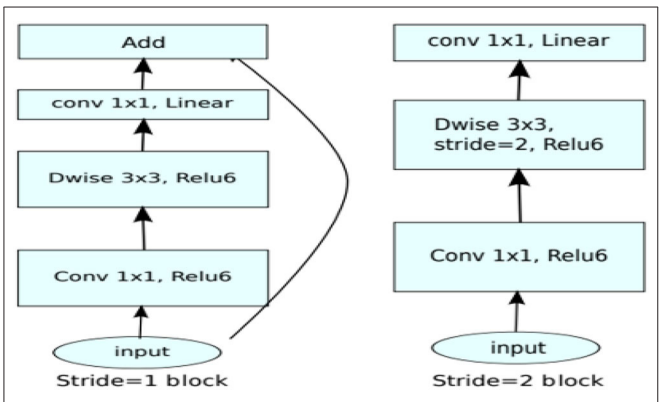


FIGURE 1 | Procedure and flow chart for Mobile-NetV2.

```
import tensorflow_hub as hub

#Use hub.KerasLayer to load a MobileNetV2 model from TensorFlow Hub
#classifier_model ="https://tfhub.dev/google/tf2-preview/mobilenet_v2/classification/4"
model = tf.keras.Sequential([
 hub.KerasLayer("https://tfhub.dev/google/tf2-preview/mobilenet_v2/feature_vector/4",
 output_shape=[1280],
 trainable=False),
 tf.keras.layers.Dropout(0.4),
 tf.keras.layers.Dense(train_generator.num_classes, activation="softmax")
])
model.build([None, 224, 224, 3])
model.summary()

Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
keras_layer_1 (KerasLayer)   (None, 1280)              2257984
_________________________________________________________________
dropout_1 (Dropout)          (None, 1280)              0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 2562
=================================================================
Total params: 2,260,546
Trainable params: 2,562
Non-trainable params: 2,257,984
```

FIGURE 2 | Summary of the proposed model.

TABLE 3 | MobileNetV2 and others pre-training models.

MobileNetV2 and others pre-training models			
State-of-the-art Model	**Parameters**	**Accuracy (Percentage) %**	**Disk Space**
MobileNet	4,253,864	70–89.5	16 MB
MobileNetV2	3,538,984	71–90	14 MB
VGG16	138,357,544	71–90	528 MB
VGG19	143,667,240	71–90	549 MB
ResNet50	25,636,712	74–92	98 MB

TABLE 4 | Image data generator () parameters.

Parameter	**Value**	**Parameter**	**Value**
Rescale	*1/255*	Zca epsilon	1e-06,
Horizontal flip	True	Validation split	0.20
Zoom range	[0.5, 1.0]	Fill mode	nearest
Feature wise center	True	sample wise center	True
Rotation range	90	Width shift range	[−20,20]
Zca whitening	False	Feature wise std normalization,	False
Vertical flip	True	Sample wise std normalization	False,
Channel shift range,	0.0	Height shift range	0.5
Brightness range	[0.5, 1.0],	Shea range,	0.0

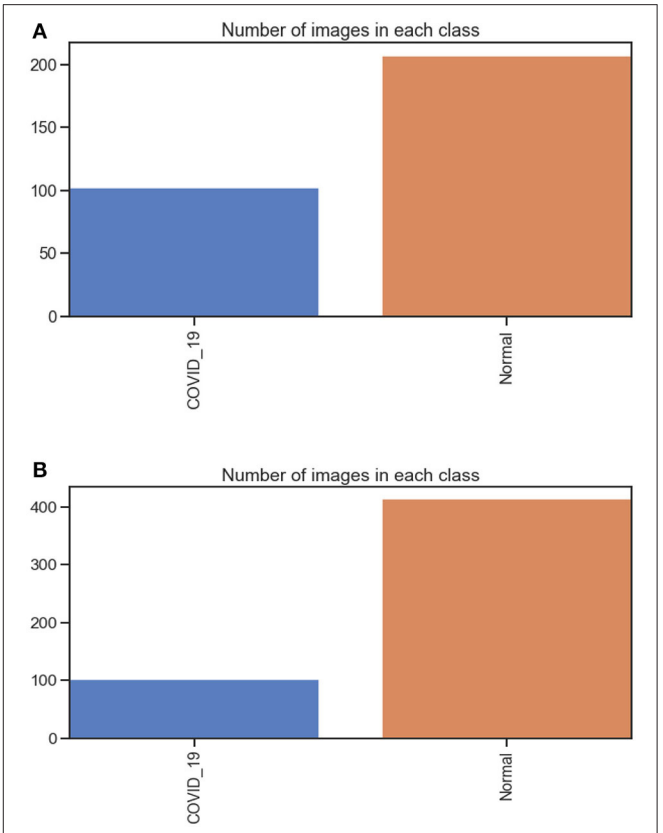


FIGURE 3 | (A) number of images in each class dataset A. **(B)** number of images in each class dataset B.

can contribute to answering this question. This article, which we contribute to, is considered to prevent the spread of the virus.

Because of the absence of a public image database of patients with COVID-19, numerous research studies reporting on the choices for the adaptive discovery of COVID-19 from X-ray images were not easy. This is because they are difficult to train, and many images are required to discover these networks without overfitting. In

this work, data augmentation of the training/testing data set was applied.

In this study, we present a COVID-19 X-ray classification technique based on transfer learning with Mobile NetV2, present in the appendix, which is a pre-trained design (without segmentation). This technique addresses the COVID-19 image deficiency issue. As an alternative to training the CNN from the beginning, our method fine tunes the last layer of the design pre-trained variation on Image-NET (presented in Section The Proposed Architecture net and Strategies for Using the Transfer Learning). The above methods assisted in training the versions with conveniently not easily offered images and they attained excellent efficiency.

The MobileNet model proved to be highly efficient in fruit classification with 102 varieties (41, 42) and the classification of systemic sclerosis skin (43). After reviewing the latest modern studies, the latest (30, 31) studies, and to the best of my knowledge, the MobileNet was not used to identify the coronavirus.

The main contribution of this paper is to investigate and evaluate the performance of MobileNetv2 as a lightweight

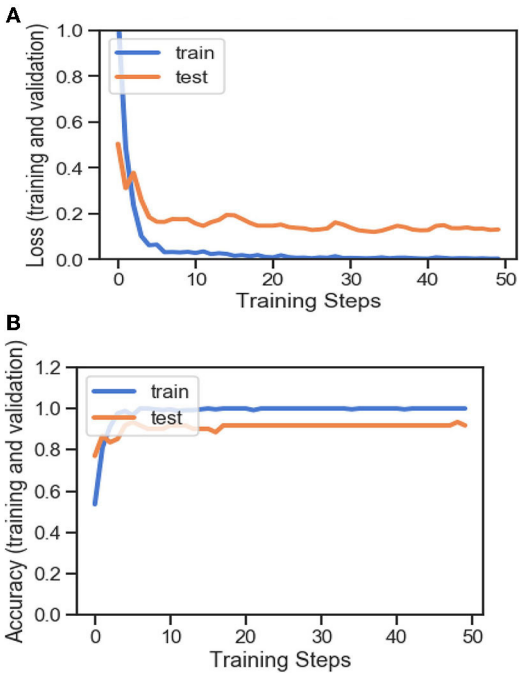


FIGURE 4 | (A) Accuracy and loss (training and validation). **(B)** Accuracy and loss (training and validation).

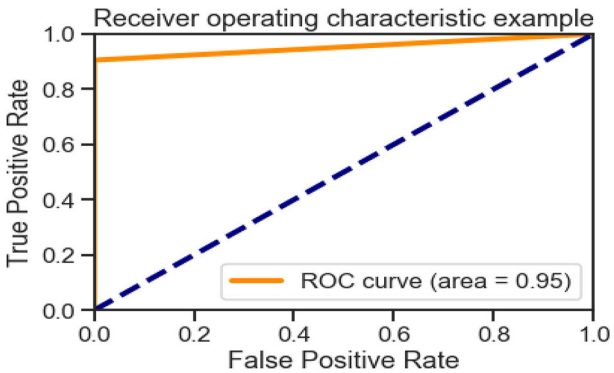


FIGURE 6 | Receiving operating characteristic (ROC) curve.

```
In [23]: predicted_ids = np.argmax(tf_model_predictions, axis=1)
         predicted_labels = dataset_labels[predicted_ids]

In [25]: from sklearn.metrics import confusion_matrix
         confusion_matrix(true_label_ids, predicted_ids)#

Out[25]: array([[19,  2],
                [ 0, 11]], dtype=int64)

In [27]: from sklearn.metrics import precision_score, recall_score
         print( "Precision:" + str(round(precision_score(true_label_ids, predicted_ids)*100,1))+"%")
         print("Recall :" + str(round(recall_score(true_label_ids,predicted_ids)*100,1))+"%")

         Precision:84.6%
         Recall :100.0%

In [28]: from sklearn.metrics import f1_score
         print("F1 score: " + str(round(f1_score(true_label_ids, predicted_ids)*100,1))+"%")

         F1 score: 91.7%
```

FIGURE 5 | Confusion matrix, precision, recall, and F1 measurements.

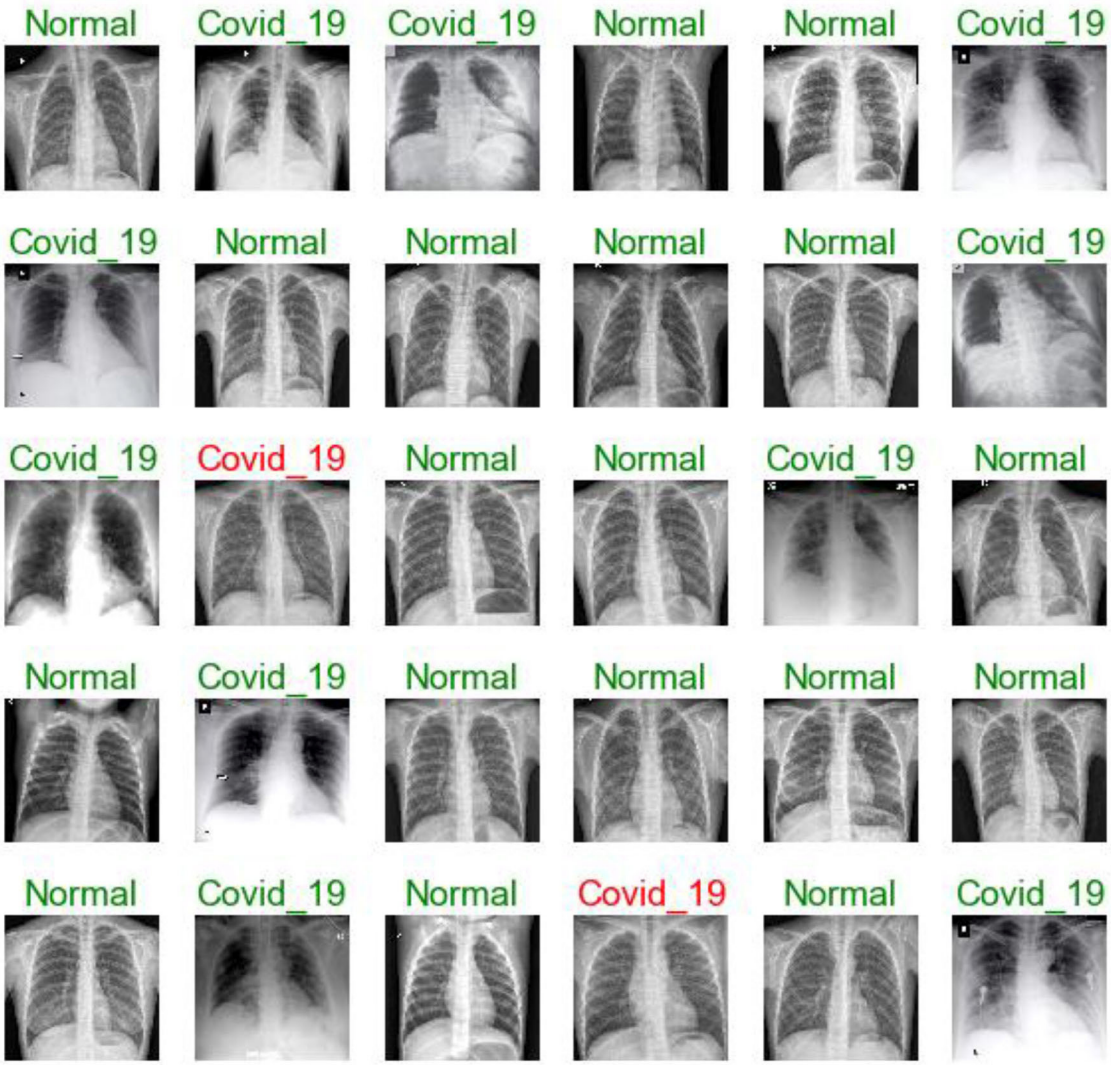


FIGURE 7 | Model predictions (green: correct, red: incorrect).

convolutional neural with transfer-learning and augmentation data for the early detection of coronavirus, to propose and adopt a new model structure for MobileNetv2 for binary classification, and to investigate the COVID-Computer-aided diagnostic (CCAD) tool based on the MobilenetV2 developed.

This paper is organized as follows: The methods and materials are described as the general projected framework in Section Materials and Methods. The results and discussions in Section Experimental Results and Discussion describe the database, experimental research studies, and measurement tools. Finally, the conclusions and future work are presented in Section Conclusion and Future Work.

MATERIALS AND METHODS

We utilized a CNN to extricate highlights from COVID-19 X-ray images. We took on an extraordinary CNN called a pre-trained show wherever the arrangement is already prepared on the Image-NET database, and it comprised numerous images (creature, plants, …on 1,000 courses). Transfer learning was utilized by exchanging weights obtained and booked right into a pre-trained adaptation system Mobile NetV2.

Transfer Learning

We had few images in the present task to conduct the training, particularly in the COVID-19 category. Thus, transfer learning

A

```
In [46]: final_loss, final_accuracy = model.evaluate(valid_generator, steps = val_steps_per_epoch)
         print("Final loss: {:.2f}".format(final_loss))
         print("Final accuracy: {:.2f}%".format(final_accuracy * 100))

         2/2 [==============================] - 4s 2s/step - loss: 0.0624 - acc: 1.0000
         Final loss: 0.06
         Final accuracy: 100.00%
```

B

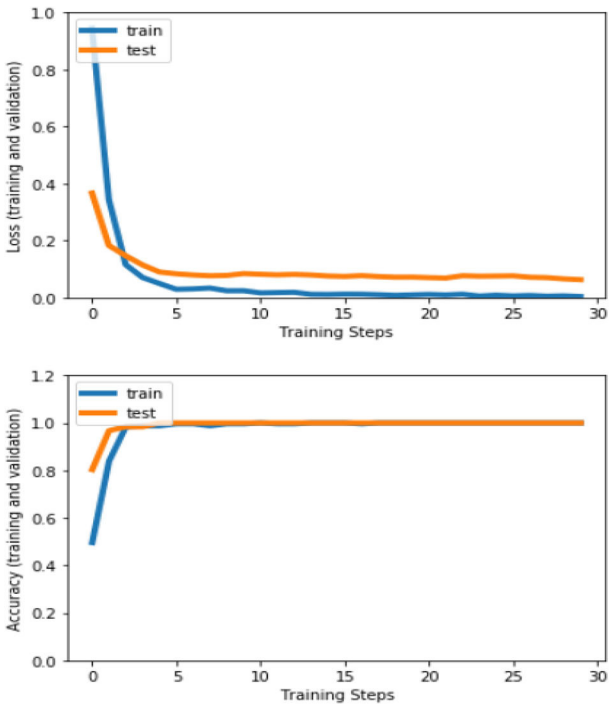


FIGURE 8 | (A) Accuracy and loss (training and validation)–Database B. **(B)** Accuracy and loss (training and validation)–Database B.

is essential. Transfer learning enhances learning in a new job by moving expertise from previous works that have already been reviewed. That is, we used techniques that have been pre-trained on image classification tasks rather than learning a new design. The advantages of transfer learning are as follows:

1. ***Higher begin:*** The first ability (before fine adjustment of the creation) on the resource design is more significant.
2. ***Higher slope:*** The rate of ability enhancement throughout the training of the source design is steeper.
3. ***Higher asymptote:*** The converged skill of the skilled version is much better. When transfer learning is applied to the design, higher precision degrees are quicker.

Description of Deep CNN

The application system is illustrated in **Tables 1**, **2**, changing the highlights from the M to N channels, with walk s and advancement perspective t. This blockage incorporates a 1 × 1 convolution layer and sometimes the depth-wise convolutional layer. In addition, in requiring the advantage of serial actuation rather than the non-linear actuation included after the pointwise convolutional layer, it fulfills its objective down-sampling by setting basis s within the depth-wise convolutional layer. The entire organized Mobile-NetV2 system (**Figure 1**) is summarized in **Table 2**. The two-dimensional (2D) convolution (conv2d) layer is the essential complication, avg-pool is the average pooling, c is the number of the result channels, and n is the copied numbers. CNNs have 19 layers; the center layers are utilized for extricating capacities and the final layers are utilized for categorization. Based on the exchange learning guideline, we utilized Mobile-NetV2 pre-trained by Image-NET as a settled separator.

The mobile-NetV2 network is a base model without top classification layers, a deal for attribute abstraction. A conv2d layer with a dropout and soft-max categorizer layer is added to the basic design in **Figure 2**.

The design of the proposed model is summarized in **Figure 2**; we notice that the total number of coefficients is 2,260,546 for the new model. Only 2,562 coefficients are trainable.

Table 3 shows the differences between the MobileNetV2 model and the other models regarding requirements and efficiency.

```
In [54]: from sklearn.metrics import confusion_matrix
         confusion_matrix(true_label_ids, predicted_ids)#

Out[54]: array([[25,  0],
                [ 0,  7]], dtype=int64)

In [55]: from sklearn.metrics import precision_score, recall_score
         print( "Precision:" + str(round(precision_score(true_label_ids, predicted_ids)*100,1))+"%")
         print("Recall :" + str(round(recall_score(true_label_ids,predicted_ids)*100,1))+"%")

         Precision:100.0%
         Recall :100.0%

In [56]: from sklearn.metrics import f1_score
         print("F1 score: " + str(round(f1_score(true_label_ids, predicted_ids)*100,1))+"%")

         F1 score: 100.0%
```

FIGURE 9 | Confusion matrix, precision, recall, and F1 measurements-Case B.

EXPERIMENTAL RESULTS AND DISCUSSION

Database Description

The first dataset includes 309 chest X-ray images (102 were infected with COVID-19 and 207 were normal) see **Figure 3**. The second dataset includes 516 chest X-rays images (102 were infected with COVID-19 and 414 were normal). All Data sets are organized in two folders, one for training and the second for testing /validation. In each folder of the two (training and testing), two open folders equal to the number of classes, then in each class, the images examples of this class are put. The dataset was obtained from two different open-source databases:

1. GitHub database [https://github (44)].
2. https://kaggle.com/c/rsna-pneumonia-detection-challenge.

They include 8,851 normal samples and 6,012 samples infected with pneumonia, and the chest X-ray image dimension is adjusted to 224 × 224 to satisfy the input conditions of the Mobile-NetV2 model.

Open-source CNN versions are readily available for training from scratch and transfer learning (customizing parts of an extant design for a new job). These versions are frequently tested on the ImageNet dataset of more than 1 million classifieds. The following section presents the proposed structure based on the MobileNetV2.

The Proposed Architecture Net and Strategies for Using the Transfer Learning

1. **Building the generator of the training data.**

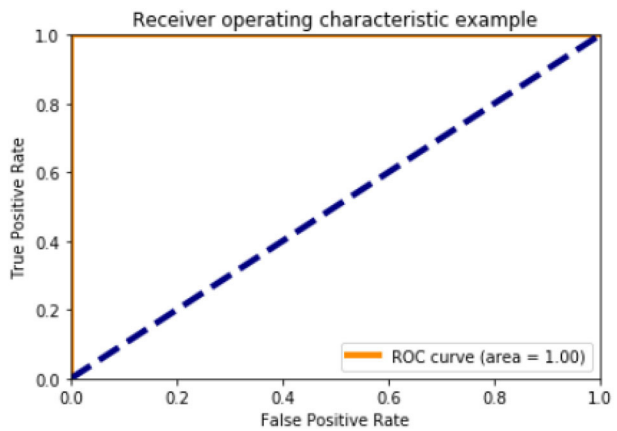


FIGURE 10 | ROC curve (Case B).

The batches of tensor image data were generated using the class ImageDataGenerator(); also, this class is used as real-time data augmentation with different parameters. With the flow_from_directory() method, which receives two arguments, the directory(from which the images are loaded for creating the batches) and target size (we Adapt the: Target size of the loaded image, which is (256, 256) by default. To (128, 128). In addition, the method receives 3 other arguments (color_mode: RGB, class_mode:' "categorical", batch_size:32).

2. **Building the validation data generator (is like building the generator of the training data).**

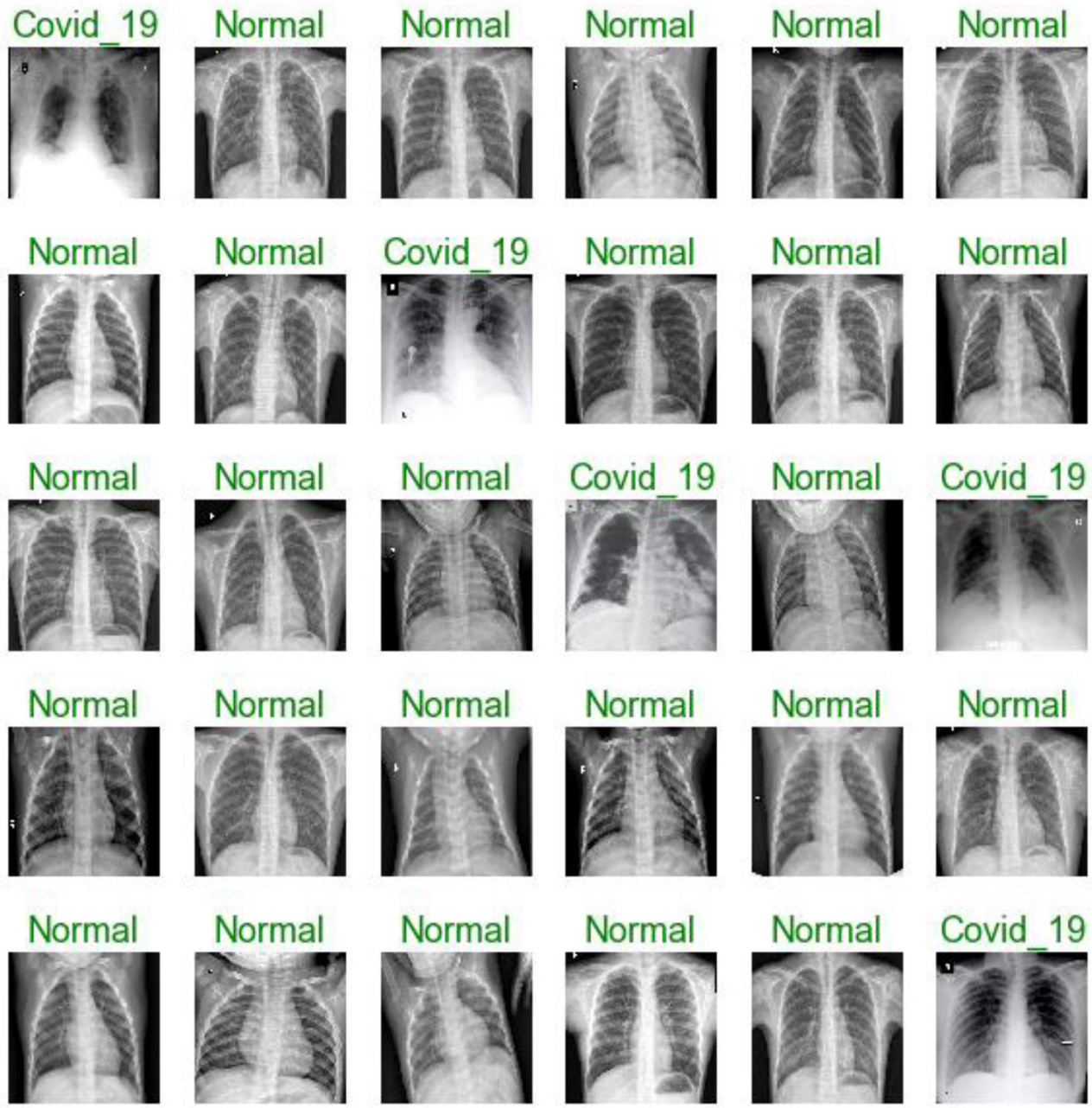


FIGURE 11 | Model predictions (green: correct, red: incorrect).

3. **Loading the MobileNetv2.**

Load the MobileNetv2 from the TensorFlow.Keras.Applications and store in the COVID19Model. Then, remove the last Final conect (FC) layers from the original model. The number of classes is 1,000, and the number of parameters is 4,231,976 trainable parameters in the network. To make this model suitable for working with our data set (Covid19) in two classes: change the size image from (224, 224, 3) to (128, 128, 3). we set in the MobileNet() function argument input_shape = (128, 128), and include_top = False (meaning load the model without FC layers) then make the model return just a vector of length equal to the number of classes. Also, the trainable argument of the loaded model is set to False.

4. **Adding new FC Layers to the modified model.**

We do not need to build a new architecture but use the architecture in the modified model stored in the COVID 19 model. Using the Sequential class in Keras, two layers were added

```
In [13]: #Use hub.KerasLayer to load a MobileNetV2 model from TensorFlow Hub
         #classifier_model ="https://tfhub.dev/google/tf2-preview/mobilenet_v2/classification/4"
         model = tf.keras.Sequential([
          hub.KerasLayer("https://tfhub.dev/google/tf2-preview/mobilenet_v2/feature_vector/4",
          output_shape=[1280],
          trainable=False),
          tf.keras.layers.Dropout(0.5),
          tf.keras.layers.Dense(train_generator.num_classes, activation="softmax")
         ])
         model.build([None, 224, 224, 3])
         model.summary()

         Model: "sequential"
         _________________________________________________________________
         Layer (type)                 Output Shape              Param #
         =================================================================
         keras_layer (KerasLayer)     (None, 1280)              2257984
         _________________________________________________________________
         dropout (Dropout)            (None, 1280)              0
         _________________________________________________________________
         dense (Dense)                (None, 2)                 2562
         =================================================================
         Total params: 2,260,546
         Trainable params: 2,562
         Non-trainable params: 2,257,984
```

FIGURE 12 | Case B with 10 epochs.

at the top of the modified Mobile Net architecture, which are the: a) Average pooling layer b) FC layer with two neurons using the Dense class. (Note that an FC layer is also called a dense layer, and therefore the class used for building the FC layer is called Dense.) Adapting two layers is better than adapting one layer. Adapting three layers is better than just adapting 1 or 2 layers.

5. **Compiling the new model.**

Use the **compile ()** method with the arguments optimizer, loss and metrics as: optimizer=tf.keras.optimizers.RMSprop (lr=0.0001), loss ="binary_crossentropy" and metrics=["accuracy"]).

6. **Finally, MobileNet transfer learning over the COVID-19 dataset.**

   The **fit–_generator()** method is used. With the argument: generator: Train generator. steps_per_epoch: set to ceil (number of training samples/batch size). epochs: Number of training epochs. validation_data: Validation generator. validation_steps: calculated as ceil (number of validation samples/batch size.

7. **Finally, save the model with extension.h5,** which is a format for saving structured data.

Data Augmentation

In machine learning (ML) and deep learning (DL) applications, the availability of large-scale, high-quality datasets plays a major role in the accuracy of the results. Keras ImageDataGenerator class provides a fast and simple method to increase your images. It offers various augmentation strategies like standardization, rotation, shifts, turns, brightness modification, etc. Data augmentation of the training/testing data set was applied to each image. For this, we use used the Keras ImageDataGenerator () class with the flowing set of constructor parameters: **Table 4**.

Experimental Environment

To survey the execution of Mobile-NetV2 with exchange learning, TensorFlow 2 (Tensorflow free) was introduced on Windows-10, a 64-bit working framework. Mobile-NetV2 was run using Python (Python free), and it was pre-trained using the Image-NET database. The exploratory framework is as follows: Intel R Center i7-8250U CPU @1.70 GHz, 1900 MHz 7 Core(s). 9 consistent Pr. and the memory was 16 GB. We used the Keras guide (45) and TensorFlow backend (46) in the implementation model.

Evaluation Metrics

In the classification Models, most researchers used only accuracy as a performance metric. But the accuracy in some cases is not enough because of the imbalance problem in the data. To measure the model's performance and avoid the problem of data imbalance in the classes, we used more than one performance metric, such as the confusion matrix, F1, Precision, Recall, and Area Under Curve (AUC), which is one of the highest used metrics for evaluation. It is used for binary classification problems. The performance of the classification models can be assessed using various metrics, such as classification precision, sensitivity, specificity, F1-score, and accuracy. Sensitivity and specificity can be calculated using the following equations (45–47): Precision: measure the true patterns correctly predicted from the total predicted patterns in a true class. Accuracy: measure the ratio of correct predictions over the total number

A

```
In [15]: import numpy as np
steps_per_epoch = np.ceil(train_generator.samples/train_generator.batch_size)
val_steps_per_epoch = np.ceil(valid_generator.samples/valid_generator.batch_size)
hist = model.fit(
 train_generator,
 epochs=10,
 verbose=1,
 steps_per_epoch=steps_per_epoch,
 validation_data=valid_generator,validation_steps=val_steps_per_epoch).history
```

```
Epoch 1/10
13/13 [==============================] - 30s 2s/step - loss: 0.5400 - acc: 0.7387 - val_loss: 0.1284 - val_acc: 0.9510
Epoch 2/10
13/13 [==============================] - 22s 2s/step - loss: 0.1965 - acc: 0.9312 - val_loss: 0.0663 - val_acc: 0.9902
Epoch 3/10
13/13 [==============================] - 31s 2s/step - loss: 0.1289 - acc: 0.9591 - val_loss: 0.0550 - val_acc: 0.9706
Epoch 4/10
13/13 [==============================] - 27s 2s/step - loss: 0.0603 - acc: 0.9778 - val_loss: 0.0425 - val_acc: 0.9804
Epoch 5/10
13/13 [==============================] - 24s 2s/step - loss: 0.0867 - acc: 0.9618 - val_loss: 0.0340 - val_acc: 0.9902
Epoch 6/10
13/13 [==============================] - 23s 2s/step - loss: 0.0510 - acc: 0.9824 - val_loss: 0.0329 - val_acc: 0.9804
Epoch 7/10
13/13 [==============================] - 23s 2s/step - loss: 0.0449 - acc: 0.9831 - val_loss: 0.0266 - val_acc: 1.0000
Epoch 8/10
13/13 [==============================] - 32s 2s/step - loss: 0.0424 - acc: 0.9817 - val_loss: 0.0264 - val_acc: 0.9902
Epoch 9/10
13/13 [==============================] - 28s 2s/step - loss: 0.0340 - acc: 0.9858 - val_loss: 0.0236 - val_acc: 0.9902
Epoch 10/10
13/13 [==============================] - 24s 2s/step - loss: 0.0282 - acc: 0.9901 - val_loss: 0.0221 - val_acc: 0.9902
```

```
4/4 [==============================] - 4s 797ms/step - loss: 0.0221 - acc: 0.9902
Final loss: 0.02
Final accuracy: 99.02%
```

B

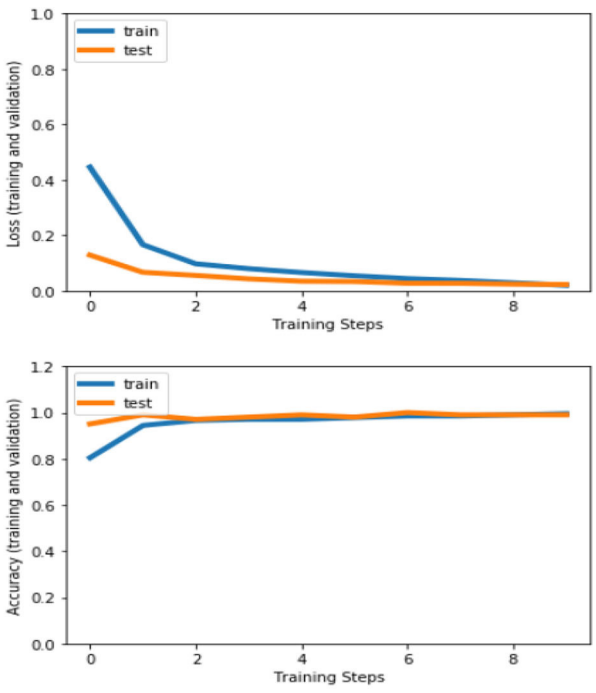


FIGURE 13 | (A) Accuracy and loss (training and validation)–Database B with 10 epochs. **(B)** Accuracy and loss (training and validation)–Database B with 10 epochs.

of examples evaluated, Recall: used to measure the fraction of positive examples that are correctly classified, and F1: a measure that provides the absolute average between accuracy and recall.

$$\text{Accuracy } (\text{for each category}) = \frac{TN + TP}{TN + TP + FN + FP},$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \text{Sensitivity} = \frac{TP}{FN + TP},$$

$$, \text{Precision} = \frac{TP}{FP + TP}, \text{and} F1 = \frac{2^{*}TP}{2^{*}TP + FP + FN}$$

where true positive (TP) is the correctly categorized number and false positive (FP) is the incorrectly recognized number. False-negative (FN) is the image of a category that is found as an additional class and true negative (TN) is an image that does not belong to any category and is not categorized as any class.

Evaluation Results

We prepared the experiment with our database utilizing the Adam optimizer and a batch measure of 30. We performed our exploration by selecting a dropout rate (0.4, 0.5). Additionally, we utilized a learning rate rise to 1e-3 for the Adam optimizer and prepared for ten epochs within the pre-training organization; in any case, we select a much lower preparation rate e^{-5} for 31 epochs during the fine-adjustment phase.

Case 1 (Dataset A)

Database B

Figures 4A,B, **8A,B**, **13A,B**, and **Table 5** represents the accuracy and loss (training and validation) for data set A, Data set B and Dataset B with only 10 epochs and a dropout of 0.5, respectively.

Figures 5, **9**, **14** and **Tables 6**, **7** show the Confusion matrix, precision, recall, and F1 measurements for data set A, Data

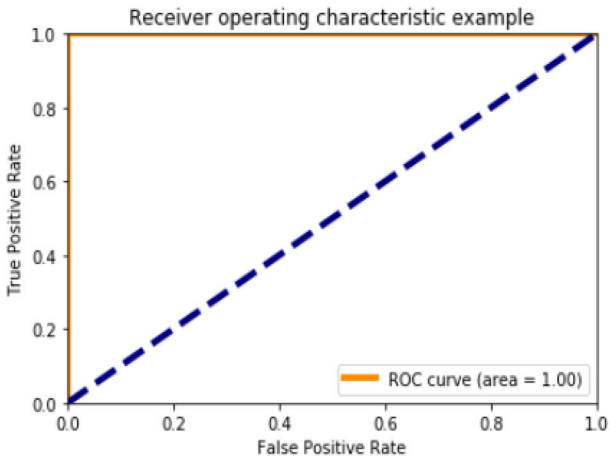


FIGURE 15 | ROC curve (Case B).

TABLE 5 | Final loss and final accuracy for all data set.

Data set	Final loss	Final accuracy
A	0.09	95.08%
B	0.06	100%
B with only 10 epochs and Dropout 0.5	0.02	99.02%

TABLE 6 | Precision, recall and F1 score for all data set.

Data set	Precision	Recall	F1 score
A	100%	100%	95.0%
B	100%	100%	100%
B with only 10 epochs and Dropout 0.5	100%	100%	100%

```
In [23]: from sklearn.metrics import confusion_matrix
         confusion_matrix(true_label_ids, predicted_ids)#

Out[23]: array([[ 8,  0],
                [ 0, 24]], dtype=int64)

In [25]: from sklearn.metrics import precision_score, recall_score
         print( "Precision:" + str(round(precision_score(true_label_ids, predicted_ids)*100,1))+"%")
         print("Recall :" + str(round(recall_score(true_label_ids,predicted_ids)*100,1))+"%")

         Precision:100.0%
         Recall :100.0%

In [26]: from sklearn.metrics import f1_score
         print("F1 score: " + str(round(f1_score(true_label_ids, predicted_ids)*100,1))+"%")

         F1 score: 100.0%
```

FIGURE 14 | Confusion matrix, precision, recall, and F1 measurements-Case B with 10 epochs.

set B, and Dataset B with only 10 epochs and a dropout of 0.5, respectively.

Figures 6, **10**, **15** and **Table 8** represents the receiver operating characteristic (ROC) Curve for data set A, Data set B, and Dataset B with only 10 epochs and a dropout of 0.5, respectively (see **Figure 12**).

Figures 7, **11**, **16** are present the model predictions results (green: correct, and red: incorrect) for data set A, Data set B, and Dataset B with only 10 epochs and a dropout of 0.5, respectively.

Based on the results (all figures and tables), it is shown that the proposed classification strategy DL MobileNetV2 might considerably influence the automated detection and extraction of essential features from X-ray images associated with the medical diagnosis of COVID-19.

The preparation of a deep CNN (D-CNN) takes days and hours, and it requires a large database to induce significant precision. Most assignments or information were related. The model coefficients learned can be utilized within the modern demonstration through transfer learning to accelerate and maximize the learning execution of the modern model. We observed from our tests that preparing D-CNNs, particularly for Mobile-NetV2, with a transfer learning strategy could hasten the preparation method.

TABLE 7 | Confusion matrix for all data set.

Data set	Confusion matrix
A	19 2 0 11
B	25 0 0 7
B with only 10 epochs and Dropout of 0.5	25 7 0 7

TABLE 8 | Roc curve area for all data sets.

Data set	Roc curve area
A	0.95
B	1.0
B with only 10 epochs and Dropout 0.5	1.0

Moreover, distinct profundity convolutions and modified remaining direct blockages can diminish the number of coefficients. Thus, Mobile-NetV2 can be rapidly sent in versatile and inserted apparatus. Because the database utilized in this study is small, we trained our model utilizing dropout, which is a proficient strategy for minimizing overfitting in neural systems by dodging sophisticated adaptations on preparing information.

DL Challenge in the Diagnosis of COVID19 Chest Xray

Some limitations of the study research can be overcome in future research, specifically, a more in-depth analysis that requires much more patient data. Establishing models to identify COVID-19 cases from similar viral cases, such as severe acute respiratory syndrome (SARS), and a higher range of common pneumonia is needed. Nonetheless, this work contributes to the possibility of an affordable, fast, and automated diagnosis of the Coronavirus disease.

In the last two years, much DL research on X-rays has been conducted to promote and propose a primary diagnostic tool for COVID-19. These studies have shown good performance results and the best DL algorithm models. These studies have proven that DL algorithms can classify or differentiate between normal and COVID19 chest X-rays. Hence, to keep the study more reliable and to be part of the primary diagnosis tool, perhaps, now is the time to move the study to classify the severity level of COVID-19 where the levels of COVID-19 severity infection are classified into five stages. They are beginning from stage 1 up to stage 5. Stage 1 is defined as the early stage of infection, where these levels classify based on SpO2 level respiratory frequency and difficulty of breathing symptoms. **Table 10** shows the severity level and its common symptoms. To certify this future works, deep learning researchers and medical doctors must work together to recognize or build the actual image samples for the algorithm.

CONCLUSION AND FUTURE WORK

This study presents a recognizable verification framework for COVID-19 based on a light neural network, Mobile-NetV2, with a transfer-learning procedure and chest X-ray images. We utilized the pre-trained neural system, Mobile-NetV2, which was prepared using the Image-NET database. The conventional

TABLE 9 | Compare the results obtained using the proposed network with other networks (37).

Network		Accuracy 2-class (%)	Sensitivity (%	Specificity (%)	F1 %
VGG19 (48)		98.75	92.85	98.75	-
MobileNet v2 (28)		97.40	99.10	97.09	-
Inception (49)		86.13	12.94	99.70*	-
Xception (50)		85.57	0.08	99.99*	-
Inception ResNet v2 (49)		84.38	0.01	99.83*	-
Proposed Network	A	93.75	100	100	95.0
(Based on MobileNetv2)	B	99.2 and 100	100	100	100

*The asterisk * denotes the measurements corresponding to sensitivity and specificity, which obtained not meaningful values due to data imbalance. In our work, Data augmentation is used in the training/testing data set using Keras ImageDataGenerator () class.*

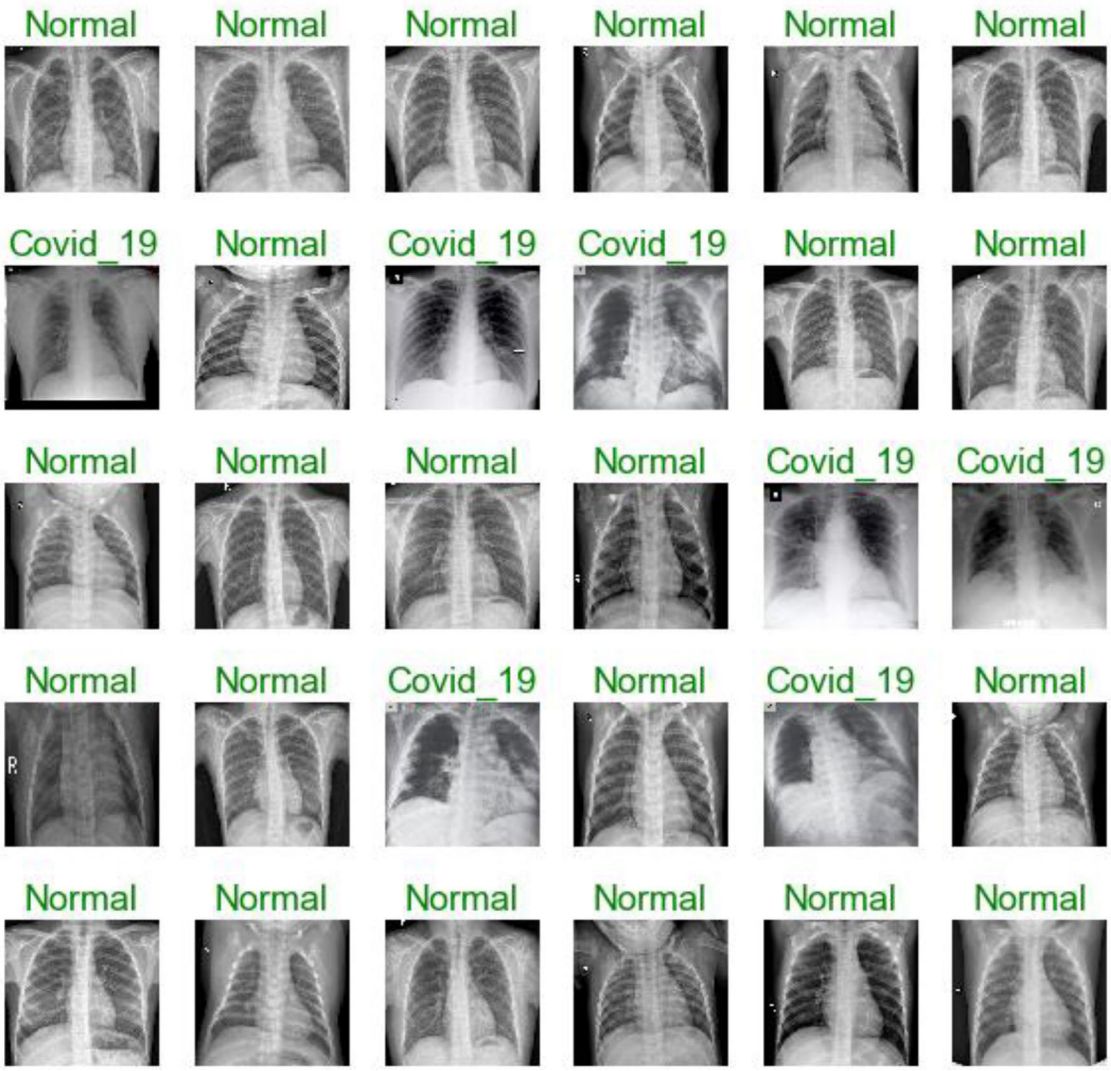


FIGURE 16 | Model predictions (green: correct, red: incorrect).

TABLE 10 | Severity level of COVID-19 (51).

Severity level	Name	Description/Individuals who
1	Asymptomatic	Have no symptoms that are consistent with COVID-19
2	Mild	Have light symptoms
3	Moderate	Show evidence of lower respiratory (SPO2 ≥94%)
4	Severe	Have SPO2 < 94%, and other conditions
5	Critical	Have respiratory failure/multiple organ dysfunction

convolution layer was used as the beat layer of a recent study. To reduce overfitting, we connected the recession to the recently included conv2d layer. Mobile-NetV2 demonstrates employment to extricate highlights, and the soft-max categorizer is employed to categorize highlights.

The performance rates indicate that computer-aided diagnostic models based on the convolutional neural MobilenetV2 networks may be employed to diagnose COVID-19. The proposed strategy achieved a distinguishing proof accuracy of 99.9 and 100% in our chest X-ray pictures database, with 309 and 512 images (102 infected with COVID-19, 207 normal; 102 infected with COVID 19, 414 normal, respectively) (see **Table 9**). The proposed framework can be employed in limited computing and low power devices such as smartphones because Mobile-NetV2 is a lightweight neural arrangement.

In the future, we plan to promote a hybrid system based on MobileNetV2 as a feature extractor with the Whale Optimization Algorithm (WOA) for efficient feature selection and other classifiers such as sport vector machine or Extra Randomized Tree (ERT) algorithm (52).

DATA AVAILABILITY STATEMENT

The datasets analysed in this study are available at: GitHub database [https://github (30)] and https://kaggle.com/c/rsna-pneumonia-detection-challenge. Further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethics approval and written informed consent were not required for this study in accordance with national guidelines and local legislation.

AUTHOR CONTRIBUTIONS

MR supervised the project. GA, HA, and SA designed the model, the computational framework, and analysed the data. AN and MR carried out the implementation. GA and HA performed the calculations. GA and SA wrote the manuscript with input from all authors. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported financially by Institution Fund projects under Grant No. (IFPRC-215-249-2020).

ACKNOWLEDGMENTS

We extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (IFPRC-215-249-2020). and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

REFERENCES

1. Lai CC, Shih TP, Ko WC, Tang HJ. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrobial Agents.* (2020) 55:105924. doi: 10.1016/j.ijantimicag.2020.105924
2. Chen W, Peter WH, Frederick GH, George FGA. novel coronavirus outbreak of global health concern. *Lancet.* (2020) 395:470–3. doi: 10.1016/S0140-6736(20)30185-9
3. Wong HY, Lam HY, Fong AH, Leung ST, Chin TW, Lo CS, et al. Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology.* (2019) 296:E72–8. doi: 10.1148/radiol.2020201160
4. Cho S, Lim S, Kim C, Wi S, Kwon T, Youn WS. Enhancement of soft-tissue contrast in cone-beam Computed tomography (CT) using an anti-scatter grid with a sparse sampling approach. *Phys Medica.* (2020) 70:1–9. doi: 10.1016/j.ejmp.2020.01.004
5. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W. Correlation of chest Computed tomography (CT) and RT-PCR testing in coronavirus disease 2019/2020 (COVID- 19) in China: a report of 1014 case. *Radiology.* (2020) 296. doi: 10.1148/radiol.2020200642
6. Kanne JP, Little BP, Chung JH, Elicker BM. Essentials for radiologists on COVID-19: an update - radiology scientific expert panel. *Radiology.* (2020) 296:E113–4. doi: 10.1148/radiol.2020200527
7. Wang L, Wong A. Covid-net, A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv:2003.09871. (2020). doi: 10.1038/s41598-020-76550-z
8. Montagnon E, Cerny M, Cadrin-Chênevert A, Hamilton V, Derennes T, Ilinca A. Deep learning work?ow in radiology: a primer. *Insights into Imag.* (2020) 11:22. doi: 10.1186/s13244-019-0832-5
9. Ibrahim A, Mohammed S, Ali HA, Hussein SE. Breast cancer segmentation from thermal images based on chaotic Salp swarm algorithm. *IEEE Access.* (2020) 8:122121–34. doi: 10.1109/ACCESS.2020.3007336
10. Han J, Zhang D, Cheng G, Liu N, Xu D. Advanced deep learning techniques for salient and category-specific object detection: a survey. *IEEE Signal Process Mag.* (2018) 35:84–100. doi: 10.1109/MSP.2017.2749125
11. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556. (2014). Available online at: http://arxiv.org/abs/1409.1556
12. Al-Dhamari A, Sudirman R, Humaimi Mahmood NH. and Transfer deep learning along with binary support vector machine for abnormal behavior detection. *IEEE Access.* (2020) 8:61085–95. doi: 10.1109/ACCESS.2020.2982906
13. Yu S, Xie L, Liu L, Xia D. Learning long-term temporal features with deep neural networks for human action recognition. *IEEE Access.* (2020) 8:1840–50. doi: 10.1109/ACCESS.2019.2962284
14. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Into Imag.* (2018) 9:611–29. doi: 10.1007/s13244-018-0639-9
15. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE.* (1998) 86:2278–324. doi: 10.1109/5.726791
16. Arayanan VB, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intel.* (2017) 39:2481–95. doi: 10.1109/TPAMI.2016.2644615
17. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inform Process Syst.* (2015) 28:91–9. Available online at: https://arxiv.org/abs/1506.01497.
18. Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: *European Conference on Computer Vision.* Springer, Cham. (2014). doi: 10.1007/978-3-319-10593-2_13
19. Minaee S, Abdolrashidi A, Su H, Bennamoun M, Zhang D. Biometric recognition using deep learning: a survey. arXiv:1912.00271. (2019).
20. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
21. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. Cnn features off-the-shelf: an astounding baseline for recognition. In: *Proceedings of the IEEE Conference on computer vision and pattern recognition workshops.* (2014). p. 806–813 doi: 10.1109/CVPRW.2014.131
22. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, et al. DeCAF: a deep convolutional activation feature for generic visual recognition. In: *International Conference on Machine Learning.* (2014). p. 647–655.
23. Nguyen L, Lin D, Lin Z, Cao J. Deep cnns for microscopic image categorization by exploiting transfer learning and feature concatenation. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS).* (2018). p. 1–5. doi: 10.1109/ISCAS.2018.8351550
24. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. (2014).
25. He K, Zhang XR. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 770–778. doi: 10.1109/CVPR.2016.90

26. Zoph B, Vasudevan V, Shlens J, Quoc Le Q. Automl for large scale image categorization and object detection. Available online at: https://kaggle.com/c/rsna-pneumonia-detection-challenge, Blog. (2017).
27. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. (2017).
28. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2015). p. 1–9. doi: 10.1109/CVPR.2015.7298594
29. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence.* (2017).
30. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc.* (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
31. Manoj M, Srivastava G, Somayaji SRK, Gadekallu TR, Maddikunta PKR, Bhattacharya S. An Incentive Based Approach for COVID-19 using Blockchain Technology. In: *IEEE Globecom Workshops.* (2020). doi: 10.1109/GCWkshps50303.2020.9367469
32. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell 1725.* (2018) 1122–31. doi: 10.1016/j.cell.2018.02.010
33. Rajaraman S, Candemir S, Kim I, Thoma G, Antani S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Appl Sci.* (2018) 8:10. doi: 10.3390/app8101715
34. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *Eur Radiol.* (2021) 31:6096–104. doi: 10.1007/s00330-021-07715-1
35. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint arXiv:2003.04655. (2020)
36. Ghoshal B, Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. arXiv preprint arXiv:2003.10769. (2020)
37. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med.* (2020) 43:635–40. doi: 10.1007/s13246-020-00865-4
38. Esmail K, Hemdan ME, Shouman MA. Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. *Compl Intell Syst.* (2021) 7:235–47. doi: 10.1007/s40747-020-00199-4
39. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed.* (2020) 196:10558. doi: 10.1016/j.cmpb.2020.105581
40. Sahiner A, Hameed AA, Jamil A. Identifying Pneumonia in SARS-CoV-2 Disease from Images using Deep Learning. *Manchester J Arti Intell Appl Sci.* (2021) 2. Available online at: https://mjaias.co.uk/mj-en.
41. Liu J, Wang X. Early recognition of tomato gray leaf spot disease based on MobileNetv2-YOLOv3 model. *Plant Methods.* (2020) 16:83. doi: 10.1186/s13007-020-00624-2
42. Xiang Q, Wang X, Li R, Zhang G, Lai J, Hu Q. Fruit image classification based on Mobilenetv2 with transfer learning technique. In: *Proceedings of the 3rd International Conference on Computer Science and Application Engineering.* (2019). doi: 10.1145/3331453.3361658
43. Akay M, Du Y, Sershen CL, Wu M, Chen TY, Assassi S, et al. Deep learning classification of systemic sclerosis skin using the MobileNetV2 model. *IEEE Eng Med Biol.* (2021) 2:104–10. doi: 10.1109/OJEMB.2021.3066097
44. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988. (2020)
45. Chollet FK. *A Python Deep Learning Library*. Available online at: https://keras.io (2015). p. 48.
46. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467. (2015). Available online at: https://www.tensorflow.org/api_docs/python/tf/keras/backend
47. Abu-Zinadah H, Abdel Azim G. Segmentation of epithelial human type 2 cells images for the indirect immune fluorescence based on modified quantum entropy. *EURASIP J Image Video Process.* (2021). 2021:1–9. doi: 10.1186/s13640-021-00554-6
48. Ragab M, Eljaaly K, Alhakamy NA, Alhadrami HA, Bahaddad AA, Abo-Dahab SM, et al. Deep ensemble model for COVID-19 diagnosis and classification using chest CT images. *Biology.* (2022) 11:43. doi: 10.3390/biology11010043
49. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. (2015)
50. Chollet FXD. learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2017). p. 1251–1258. doi: 10.1109/CVPR.2017.195
51. Markom MA, Taha SM, Adom AH, Sukor AA, Nasir AA, Yazid H. A review: deep learning classification performance of normal and COVID-19 chest X-ray images. *J Phys.* (2021) 2071:012003. doi: 10.1088/1742-6596/2071/1/012003
52. Abbas S. BCD-WERT: A novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *Peer J Comput Sci.* (2021) 7:e390. doi: 10.7717/peerj-cs.390

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ragab, Alshehri, Azim, Aldawsari, Noor, Alyami and Abdelkhalek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2022.835938

Evaluation of the Predictors for Unfavorable Clinical Outcomes of Degenerative Lumbar Spondylolisthesis After Lumbar Interbody Fusion Using Machine Learning

Shengtao Dong[1,2†], *Yinghui Zhu*[1,3†], *Hua Yang*[4†], *Ningyu Tang*[1], *Guangyi Huang*[2], *Jie Li*[1*] *and Kang Tian*[1*]

[1] Department of Bone and Joint, First Affiliated Hospital, Dalian Medical University, Dalian, China, [2] Department of Spine Surgery, Second Affiliated Hospital of Dalian Medical University, Dalian, China, [3] Department of Orthopedics, Dalian No. 3 People's Hospital, Dalian, China, [4] Department of Otolaryngology, Head and Neck Surgery, Second Affiliated Hospital of Dalian Medical University, Dalian, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Praveen Kumar,
VIT University, India
Yan-Hua Chen,
Wuhan University, China

***Correspondence:**
Jie Li
lijie7107@163.com
Kang Tian
dmu-tiankang@outlook.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 15 December 2021
Accepted: 06 January 2022
Published: 03 March 2022

Citation:
Dong S, Zhu Y, Yang H, Tang N, Huang G, Li J and Tian K (2022) Evaluation of the Predictors for Unfavorable Clinical Outcomes of Degenerative Lumbar Spondylolisthesis After Lumbar Interbody Fusion Using Machine Learning. Front. Public Health 10:835938. doi: 10.3389/fpubh.2022.835938

Background: An increasing number of geriatric patients are suffering from degenerative lumbar spondylolisthesis (DLS) and need a lumbar interbody fusion (LIF) operation to alleviate the symptoms. Our study was performed aiming to determine the predictors that contributed to unfavorable clinical efficacy among patients with DLS after LIF according to the support vector machine (SVM) algorithm.

Methods: A total of 157 patients with single-segment DLS were recruited and performed LIF in our hospital from January 1, 2015 to October 1, 2020. Postoperative functional evaluation, including ODI and VAS were, performed, and endpoint events were defined as significant relief of symptom in the short term (2 weeks postoperatively) and long term (1 year postoperatively). General patient information and radiological data were selected and analyzed for statistical relationships with the endpoint events. The SVM method was used to establish the predictive model.

Results: Among the 157 consecutive patients, a postoperative unfavorable clinical outcome was reported in 26 patients (16.6%) for a short-term cohort and nine patients (5.7%) for a long-term cohort. Based on univariate and multivariate regression analysis, increased disc height (DH), enlarged facet angle (FA), and raised lateral listhesis (LLS) grade were confirmed as the risk factors that hindered patients' short-term functional recovery. Furthermore, long-term functional recovery was significantly associated with DH alone. In combination with the SVM method, a prediction model with consistent and superior predictive performance was achieved with average and maximum areas under the receiver operating characteristic curve (AUC) of 0.88 and 0.96 in the short-term

cohort, and 0.78 and 0.82 in the long-term cohort. The classification results of the discriminant analysis were demonstrated by the confusion matrix.

Conclusions: The proposed SVM model indicated that DH, FA, and LLS were statistically associated with a clinical outcome of DLS. These results may provide optimized clinical strategy for treatment of DLS.

Keywords: degenerative lumbar spondylolisthesis, clinical outcome, machine learning, predictor, lumbar interbody fusion

INTRODUCTION

As the aging society comes, degenerative lumbar spondylolisthesis (DLS) with symptomatic stenosis presents a serious challenge to healthcare (1, 2). After struggling with the dilemma of decompress treatment, spine surgeons have recognized that it will cause inferior postoperative outcomes without fusion, including persistent pain, pseudarthrosis, and progressive degenerative changes (3–5). Therefore, lumbar interbody fusion (LIF), in combination with decompression, has become the main surgical options for the treatment of DLS.

We noted that the negative feedback from the patients is the main reason for the increased reoperation rate and decreased treatment satisfaction, which result in contradictions or bothers between doctors and patients and heavy substantial burden on the healthcare systems (6).

The current dominant functional scales for evaluating symptoms of degenerative spinal diseases are the Visual Analog Scale (VAS) and the Oswestry Disability Index (ODI) (1, 7–9). However, there are few studies on surgical efficacy evaluation based on patient-reported outcomes (PROs) at present (7).

Machine learning (ML) has proved to be an effective data-driven automatic modeling mechanism in processing numerous problems in the biomedical field (10–13). Support vector machine (SVM) is a well-established classifier, which was applied for accurately patients' classification and treatment options improvement (14).

Considering there is a lack of superior predictive models, the purpose of this study was to investigate the clinical distribution of PROs and to establish a predictive model to evaluate the risk of unfavorable clinical outcomes in short- and long-term postoperative cohorts using the SVM algorithm.

METHODS

Study Design and Participants

The diagnosis of DLS required a minimum 3-mm vertebral displacement observed in lateral X-rays, and patients were selected from January 1, 2015 to October 1, 2020 at our institution. Inclusion criteria: (1) age ranged from 50 to 80 years old, (2) patients diagnosed with symptomatic DLS, (3) single-level spondylolisthesis and LIF, (4) traceability of clinical parameters, radiological variables, and surgical records, (5) at least 1 year of follow-up available. Exclusion criteria: (1) American Society of Anesthesiologists (ASA) rating of III or higher, (2) multi-level decompression and LIF with pedicle screws, (3) previous lumbar surgical intervention, (4) revision surgery, (5) spinal tumor, deformity, fracture, infection or other spine diseases, and (6) rejection of implants.

The Ethics Committee of our hospital approved this study. All data were extracted from hospital electronic system.

Data Collection

General patient information was recorded, including age, sex, body mass index (BMI), smoking, diabetes, Oswestry Disability Index (ODI), Visual Analog Scale (VAS), fusion technique, operation level, bone graft methods, cement augmentation, and surgical complications.

All patients received a complete imaging evaluation, including standard anteroposterior and lateral X-rays, CT, dual-energy X-ray (DEXR), and MRI. Based on X-rays results, we recorded and analyzed the following radiological variables: sacral slope (SS), the angle between the sacral plate and the horizontal axis. Draw the line connecting the midpoint of the femoral head to the midpoint of the sacral plate, which formed an angle with the vertical line of the sacral plate as pelvic incidence (PI) and an angle with the longitudinal axis as pelvic tilt (PT) (**Figure 1A**). Slip degree (SD), the horizontal distance of posterior wall between the slipped and upper vertebrae; lumbar lordosis (LL), the angle between the upper endplate of L1 and the upper endplate of S1; segment lordosis (SL), the angle between the upper endplate of the upper vertebrae and the upper endplate of the lower vertebrae of the slipped segment; disc height (DH), the average of anterior disc height and a posterior disc height (**Figure 1B**); lateral listhesis (LLS) was defined the distance between vertebral bodies (**Figure 1C**). Insertion depth, as 60% of the vertebral depth, was treated as the minimum screw purchase at our institution; the patients were assigned to the S group and the L group (**Figure 2**). The facet angle (FA) between the facet line and the line connecting the bilateral dorsal points was measured on CT (**Figure 3**). Osteoporosis was defined as T-score < -2.5 SD on dual-energy X-ray absorptiometry.

PROs were determined as function evaluations. ODI was used to describe the patient's quality of life; a smaller value

Abbreviations: DLS, Degenerative lumbar spondylolisthesis; LIF, Lumbar interbody fusion; SVM, Support vector machine; ASA, American Society of Anesthesiologists; BMI, Body mass index; LLS, Lateral listhesis; ODI, Oswestry Disability Index; VAS, Visual Analog Scale; PROs; Patient-reported outcomes; CT, Computed tomography; DEXR, Dual-energy X-ray; MRI, Magnetic resonance imaging; SD; Slip degree; LL, Lumbar lordosis; SL, Segment lordosis; SS, Sacral slope; PI, Pelvic incidence; PT, Pelvic tilt; DH, Disc height; FA, Facet angle; TLIF, Transforaminal lumbar interbody fusion; PLIF, Posterior lumbar interbody fusion; PMMA, Polymethylmethacrylate.

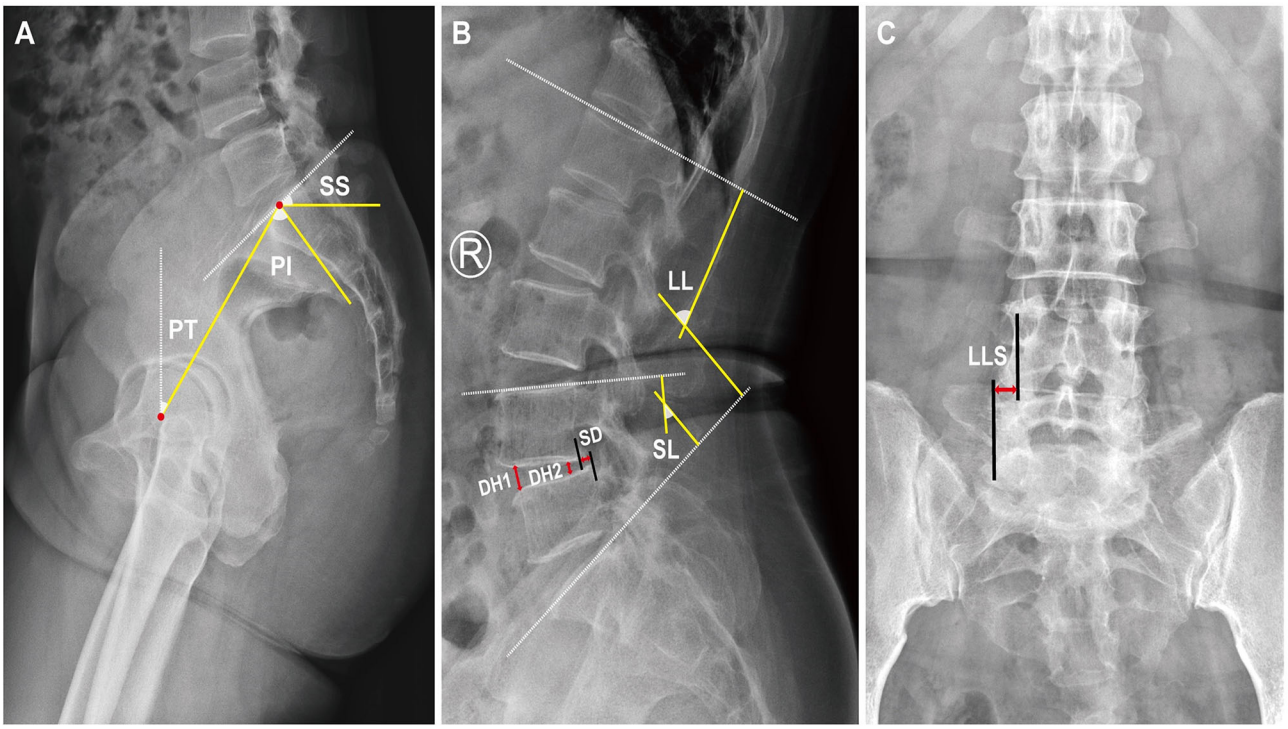


FIGURE 1 | A schematic diagram of spinopelvic parameters, lumbar parameters, and lateral listhesis (LLS). Spinopelvic parameters **(A)** included pelvic incidence (PI), pelvic tilt (PT), and sacral slope (SS). Lumbar parameters **(B)** included lumbar lordosis (LL), segment lordosis (SL), slip degree (SD), and disc height (DH). Lateral listhesis (LLS) was measured according to **(C)**.

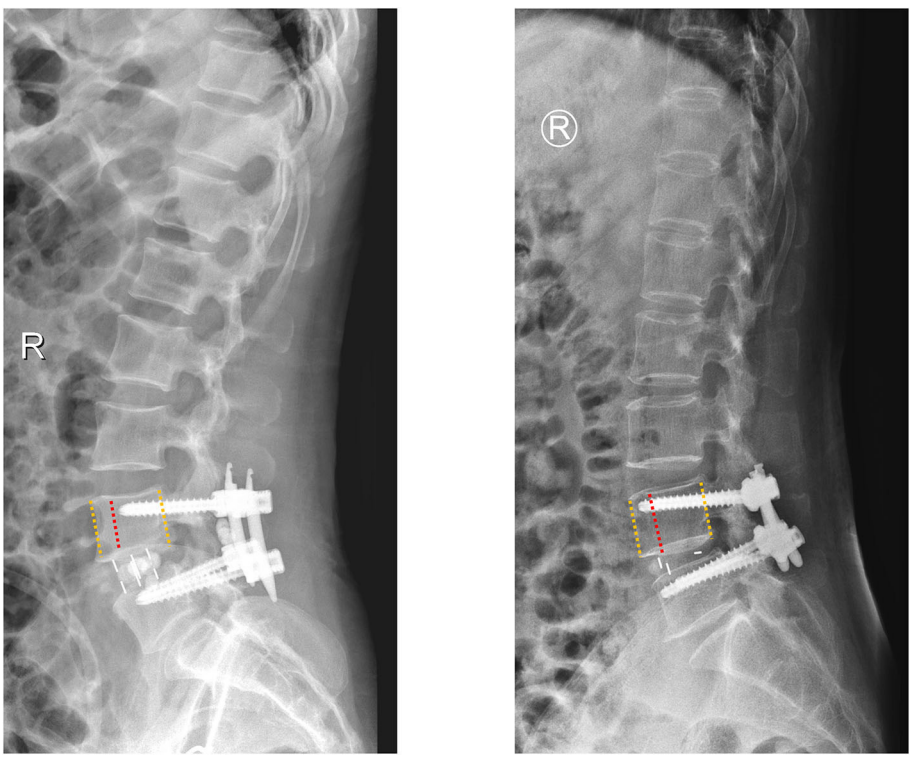


FIGURE 2 | A schematic diagram of insertion depth; the patients were assigned to the S group (the screws accounted for 60–80% of the anteroposterior diameter of vertebral body) and the L group (the screws accounted for more than 80% of the anteroposterior diameter of vertebral body).

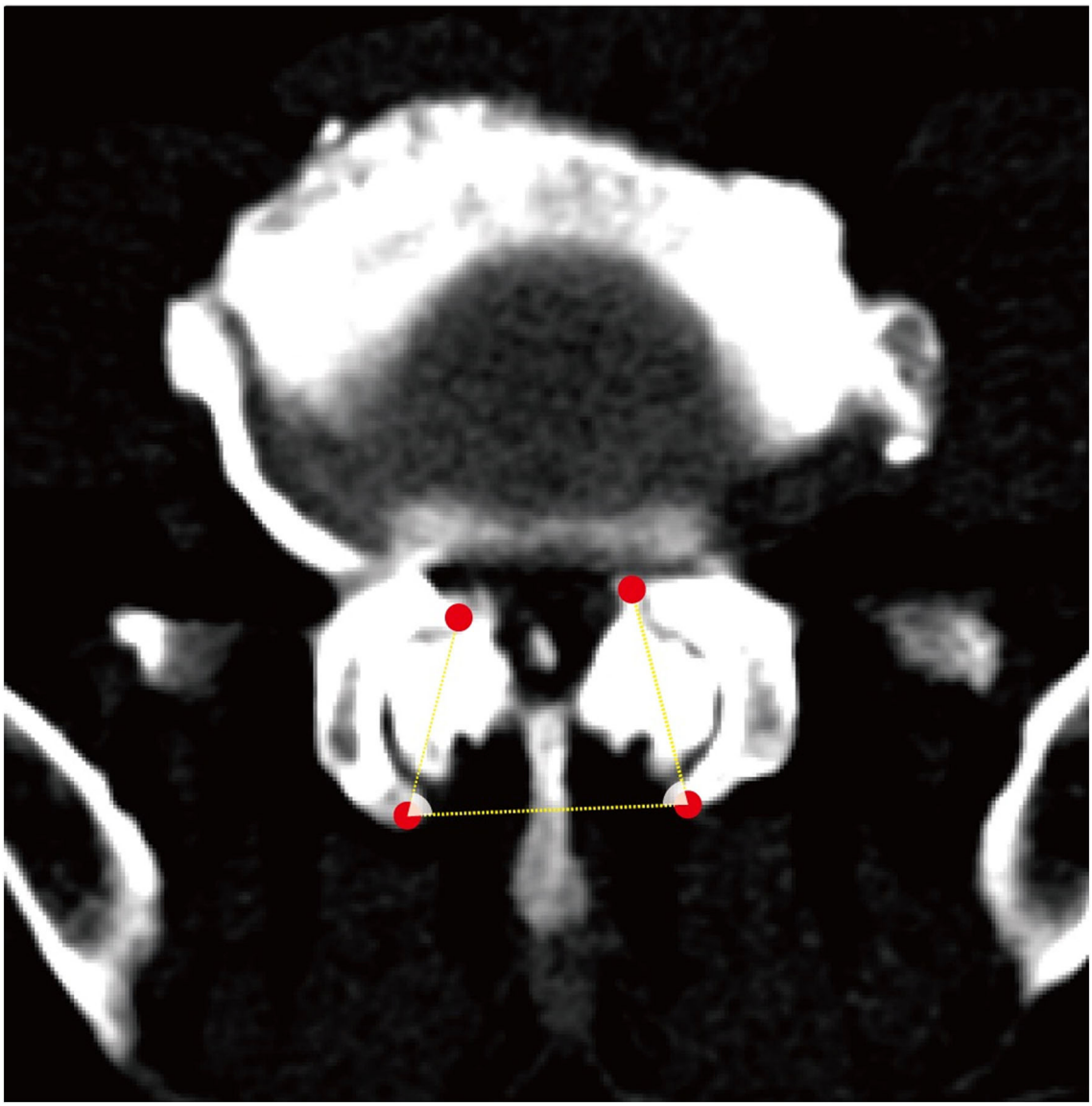


FIGURE 3 | A schematic diagram of the facet angle. A facet line was drawn between the 2 end points of each facet. The angle was measured between the facet line and the line connecting the bilateral dorsal points and recorded the average value.

suggested better function. VAS was chosen to measure the patient's back pain; a decreased value suggested reduced pain. Based on practical experience and literature review, significant symptomatic improvement was observed in patients with lumbar spondylolisthesis after decompression and fusion. (7) We defined the short-term endpoint event as a 60% improvement in VAS and ODI 2 weeks postoperatively, and long-term endpoint event as a 90% improvement in VAS and ODI 1 year postoperatively.

Surgical Technique

All recruited patients were operated by the same surgeon. Prone position after general anesthesia has taken effect. The transforaminal lumbar interbody fusion (TLIF) group and posterior lumbar interbody fusion (PLIF) group designed their individual landmarks separately with the assistance of C-arm fluoroscopy. The TLIF group used a percutaneous fixation approach to implant screws *via* the pedicle, and PLIF used a conventional open approach. If the primary surgeon felt it necessary (preoperative diagnosis of osteoporosis or intraoperative finding of bone weakness), the tip of the vertebroplasty needle was inserted into the central point of the vertebral body under fluoroscopic guidance, and, approximately, 2–3 ml of polymethylmethacrylate (PMMA) bone cement was injected. The corresponding lamina is removed and retained for bone grafting, and the upper and lower cartilage endplates of the intervertebral disc are scraped with a ring curette after debridement of the annulus fibrosus and nucleus pulposus. The rods are inserted bilaterally, and traction is applied to

lift the slipped segment and ensure uniformity with the upper and lower vertebrae as much as possible. Some patients might receive cement augmentation due to screw retraction at this stage. After fluoroscopic demonstration of satisfactory correction of the slippage, an appropriately sized cage is filled with allograft and/or autologous bone, and the processed interbody fusion cage is inserted into the interbody space to support the vertebral body (**Figure 4**).

Support Vector Machine

VAPNIK first conceived the SVM technique as a classification method for both linear and non-linear data two decades ago. The input vectors were mapped to a high-dimensional feature space *via* preselected non-linear mappings, and the optimal classification hyperplane was constructed in this space. The support variables eventually obtained are exactly the nearest sample points. The data input format was recorded as "csv" or "xlxs" and analyzed depending on the Python programming language (Python 3.8.0, Python Software Foundation).

Statistical Analysis

For continuous variables, normal distribution was subject to Student's *T*-tests, and non-normal distribution was subject to Mann–Whitney test. Chi-square test was adopted to analyze categorical variables. Continuous variables and categorical variables are presented as means ± standard deviations (SD) and relative frequencies and percentages, respectively. In addition, we applied logistic regression analysis to identify the predictors of unfavorable clinical outcomes within 2 weeks postoperatively. The statistically significant variables obtained from the univariate regression analysis were again subjected to multivariate regression analysis. Furthermore, the selected variables were put into the SVM classifier. Statistical differences were defined as $p < 0.05$. All statistical analysis was performed with SPSS version 22.0 (IBM SPSS, Armonk, New York).

RESULTS

From January 2015 to October 2020, a total of 157 consecutive patients met the inclusion criteria after excluding three patients due to follow-up loss. The demonstrative information of the patients was observed in **Table 1**. The mean age of the patients was 63.84, and the majority was females (approximately two times of males). Based on the selection basis of degenerative disease, 43 patients (27.39%) were diagnosed with osteoporosis, and 52 patients (33.12%) received intraoperative cemented augmentation. All the patients underwent single-segment fusion, with the primary level of fusion at L4/5 and TLIF as the dominant fusion approach. Autograft bone was the preferred bone graft method. Two patients had intraoperative screw retraction, and subsequent short-term poor functional recovery was observed during follow-up. The fusion rate was 95.54%, and, among the seven patients who achieved inadequate fusion, three exhibited consistent low back pain and one underwent revision surgery as the result. According to the PROs, 26 patients and nine patients were eventually assigned to the short- and long-term unfavorable cohorts; the score summary is shown in **Table 2**. Serious complications, such as neurological deficits, deep vein thrombosis, and wound infection, did not occur in this series of patients. However, we identified two patients with postoperative urinary retention and one patient who underwent revision surgery for persistent low back pain.

Table 3 compared the preoperative and postoperative radiological parameters. There were no statistically significant differences both in the short- and long-term cohorts, including

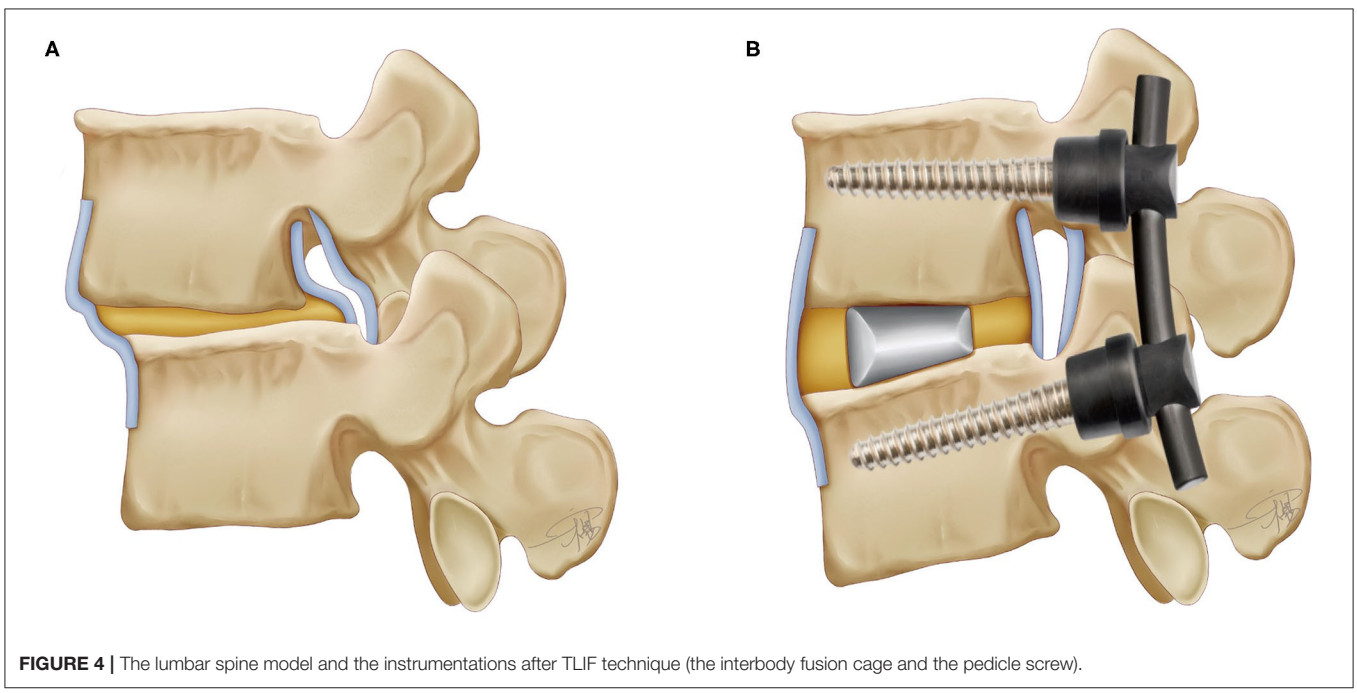


FIGURE 4 | The lumbar spine model and the instrumentations after TLIF technique (the interbody fusion cage and the pedicle screw).

TABLE 1 | Demonstrative information of the patient cohort.

Variables	Patients (*N* = 157)
Age (years), mean ± SD	63.84 ± 8.23
Sex (male/female), *n* (%)	52 (33.12%)/ 105 (66.88%)
BMI (kg/m^2), mean ± SD	24.39 ± 2.32
Smoking, *n* (%)	23 (14.65%)
Diabetes mellitus, *n* (%)	26 (16.56%)
Osteoporosis, *n* (%)	43 (27.39%)
Fusion level, *n* (%)	
L3/4	18 (11.46%)
L4/5	75 (47.77%)
L5/S1	64 (40.77%)
Fusion approach, *n* (%)	
TLIF	82 (52.23%)
PLIF	52 (33.12%)
TLIF + PLIF	23 (14.65%)
Types of bone graft	
Autograft bone	78 (49.68%)
Allograft bone	41 (26.12%)
Autograft + allograft bones	38 (24.20%)
Cement augmentation, *n* (%)	52 (33.12%)
Screw retraction, *n* (%)	2 (1.27%)
Fusion rate (fused/unfused), *n* (%)	150 (95.54%)/ 7 (4.46%)
Short-term unfavorable cohort, *n* (%)	26 (16.56%)
Long-term unfavorable cohort, *n* (%)	9 (5.73%)

TABLE 2 | Description of changes in DOI and VAS scores.

Oswestry disability index and the visual analog scale	Pre-operation	Two weeks after operation (F/UF)	One year after operation (F/UF)
ODI score	63.93 ± 10.35	28.95 ± 10.01 (26.89/ 39.35)	8.17 ± 3.20 (7.89 /12.78)
VAS score	6.96 ± 1.20	2.78 ± 1.26 (2.49/ 4.23)	0.69 ± 0.67 (0.64 /1.44)

F, favorable clinical outcomes; UF, unfavorable clinical outcomes.

SD, LL, and PI. In the present study, we found that DH, LLS, PT, and FA were significantly higher in the long-term unfavorable cohort than in the counterpart, and a higher proportion in the short screw group. In evaluating the long-term clinical efficacy, the differences of DH, LLS, and FA in the unfavorable group were statistically significant.

Univariate and multivariate regression analyses indicated that preoperative LLS ($p = 0.005$), postoperative DH ($p = 0.004$), and FA ($p = 0.030$) were independent risk factors in a short-term adverse outcome (**Table 4**), while only postoperative DH ($p = 0.038$) was independently associated with a long-term outcome (**Table 5**).

To predict the risk of an adverse outcome in patients with spondylolisthesis, we developed SVM models based on the above-mentioned independent predictors. We evaluated the ROC curve of the model and calculated the area under the ROC curve. Furthermore, 10-fold cross-validation was performed to evaluate the predictive power of the algorithm. For patients who fed back adverse PROs in the short term, the SVM model showed satisfactory classification performance with mean and maximum AUC values of 0.88 and 0.96, respectively (**Figure 5**). Among all routine indicators, the only predictor of a long-term outcome (1 year) was postoperative DH, and the long-term SVM model showed slight weakening of performance compared to the short-term SVM, with mean and maximum AUC values of 0.78 and 0.82, respectively (**Figure 6**). Based on the confusion matrix, these indices provide evidence for performance evaluation, and the following equations are presented:

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

The confusion matrix showed the models were excellent at predicting the risk of unfavorable clinical outcomes in patients with spondylolisthesis with the ACC of 93.7% and precision of 88.5% in the short-term cohort and the ACC of 86.7% and precision of 66.7% in the long-term cohort.

DISCUSSION

No consensus has been achieved on whether to perform interbody fusion for degenerative lumbar spondylolisthesis. Chan et al. (15) concluded that their evidence that supported decompression alone was a more beneficial intervention to minimize surgery-related trauma and postoperative morbidity, and had comparable efficacy to additional interbody fusion, especially for low-grade spondylolisthesis. Ghogawala et al. controverted this conclusion (5). After 4-year follow-up, they found decompression with lumbar fusion was associated with a significant improvement in physical health-related quality of life compared to laminectomy alone. The abnormal sagittal alignment of the lumbar spine can lead to increased energy demands in non-resting positions, resulting in early fatigue and exercise intolerance (9). Therefore, the aim of DLS surgery is to restore the normal anatomical sequence and stability of the lumbar spine based on the removal of spinal stenosis (7). In the present study, all the patients underwent decompressive laminectomy and instrumented lumbar spinal fusion to achieve the above-mentioned treatment goals. PROs were direct and convenient evidence to measure the quality of DLS surgeries. Meanwhile, the VAS and ODI scores were regarded as the primary evaluation tools; these subjective data offered "real world" insights into guiding and improving surgical techniques (15, 16). Moreover, a series of reported model evaluation metrics indicate that the deep learning can provide satisfactory performance in distinguishing target patients, demonstrating good clinical application (17, 18). A broad application prospect requires additional data science techniques to support, including supervised dimensionality reduction for big data (19), deep learning networks for diabetic retinopathy (20), and identification of malnourished patients (21). Using

TABLE 3 | Comparison of the radiologic variables.

Variable	Total ($n = 157$)	Short-term cohort			Long-term cohort		
		Favorable group ($n = 131$)	Unfavorable group ($n = 26$)	*p*	Favorable group ($n = 148$)	Unfavorable group ($n = 9$)	*p*
DH (mm)							
Preop	8.60 ± 2.80	8.40 ± 2.80	9.60 ± 2.63	0.045	8.57 ± 2.81	9.00 ± 2.71	0.659
Postop	5.70 ± 1.84	5.22 ± 1.43	8.11 ± 1.75	<0.001	5.57 ± 1.79	7.84 ± 1.14	<0.001
LLS (mm)							
Preop	7.67 ± 3.09	6.83 ± 2.01	11.89 ± 4.02	<0.001	7.43 ± 2.91	11.56 ± 3.50	<0.001
Postop	1.98 ± 0.73	1.93 ± 0.63	2.24 ± 1.09	0.172	1.95 ± 0.70	2.53 ± 0.84	0.021
SD (mm)							
Preop	17.76 ± 2.85	17.72 ± 2.86	17.94 ± 2.74	0.733	17.77 ± 2.87	17.51 ± 2.37	0.789
Postop	6.86 ± 1.70	6.83 ± 1.72	7.00 ± 1.61	0.629	6.82 ± 1.68	7.40 ± 1.89	0.326
LL (°)							
Preop	38.87 ± 2.25	38.83 ± 2.26	39.10 ± 2.17	0.581	38.92 ± 2.24	38.16 ± 2.27	0.327
Postop	45.26 ± 2.52	45.34 ± 2.51	44.89 ± 2.54	0.413	45.30 ± 2.53	44.62 ± 2.33	0.435
SL (°)							
Preop	16.40 ± 2.79	16.43 ± 2.87	16.24 ± 2.37	0.729	16.38 ± 2.82	16.66 ± 2.26	0.778
Postop	24.64 ±1.65	24.77 ± 1.67	24.01 ± 1.38	0.020	24.67 ± 1.67	24.28 ± 1.13	0.501
SS (°)							
Preop	36.10 ± 2.30	36.20 ± 2.26	35.61 ± 2.42	0.234	36.10 ± 2.29	36.13 ± 2.37	0.973
Postop	37.81 ± 2.83	38.08 ± 2.78	36.48 ± 2.72	0.008	37.87 ± 2.81	36.83 ± 3.04	0.286
PI (°)							
Preop	54.14 ± 2.57	54.31 ± 2.54	53.30 ± 2.56	0.068	54.22 ± 2.58	52.78 ± 1.95	0.104
Postop	52.87 ± 2.39	52.94 ± 2.41	52.49 ± 2.27	0.377	52.91 ± 2.41	52.18 ± 1.94	0.377
PT (°)							
Preop	18.03 ± 2.76	17.97 ± 2.77	18.31 ± 2.69	0.572	18.03 ± 2.75	18.06 ± 3.04	0.977
Postop	15.17 ± 1.44	15.06 ± 1.43	15.78 ± 1.37	0.020	15.14 ± 1.43	15.76 ± 1.49	0.208
Facet angle (°)	57.71 ± 7.29	55.15 ± 6.07	64.59 ± 7.79	<0.001	56.29 ± 6.95	63.65 ± 9.01	0.003
Insertion depth (%)				0.010			0.627
S group	103 (65.61%)	80 (61.07%)	23 (88.46%)		96 (64.87%)	7 (77.78%)	
L group	54 (34.39%)	51 (38.93%)	3 (11.54%)		52 (35.13%)	2 (22.22%)	

Preop, preoperative; Postop, postoperative; DH, disc height; LLS, lateral listhesis; SD, slip degree; LL, lumbar lordosis; SL, segment lordosis; SS, sacral slope; PI, pelvic incidence; PT, pelvic tilt; S group, the screws account for 60–80% of the anteroposterior diameter of vertebral body; L group, the screws accounted for more than 80% of the anteroposterior diameter of vertebral body.

the SVM model, postoperative DH, FA, and preoperative LLS were eventually identified as risk factors in unfavorable clinical outcomes at short-term follow-up, whereas only postoperative DH predicted clinical efficacy at long-term follow-up.

A large number of studies have demonstrated the importance of maintaining DH for achieving vertebral stability either in decompression alone or decompression with interbody fusion (3, 9). Our study indicated that the patients that attained a <6-mm DH intraoperatively were more likely to derive a sustained benefit from the decompression with a fusion procedure than the patients reporting poor clinical outcomes. The mean DH was higher in the short-term adverse group (8.11 vs. 7.84), which could be acceptably explained by progressive degeneration at the operated segment. The definition of the pathological change is adjacent segment disease (ASD), which was the condition responsible for half of the reoperations (3, 22, 23). Although PROs suggested that ASD was an unavoidable healthcare concern, we still feel encouraged by a reoperation rate of <1%. Sato et al. reported a reoperation rate of 2.2% in their 1-year follow-up (3), while this incidence reached a staggering 10% in the study by Deyo et al. (24). Furthermore, classifying patients with > 60% improvement in functional evaluation as the definitive efficacy group at 2 weeks postoperatively, we considered 6 mm as an effective threshold associated with clinical outcomes, which was consistent with previous studies (3, 25, 26). Considering the intervertebral space could serve to guide the selection of an interbody fusion cage, a careful fluoroscopic image review after restoring the slipped vertebra is necessary. Interestingly, Yen et al. reported the intradiscal vacuum phenomenon (IVP) occurring after interbody fusion was associated with increased disc height and vertebral instability (27). An intensive biomechanical analysis of this susceptible anatomy seems imperative.

The structure, locating on the posterolateral region of the vertebral body and connecting the adjacent vertebral arch, is the facet joint, which constitutes the classic triple-joint complex with

TABLE 4 | Univariate and multivariate logistic regression analysis of unfavorable clinical outcomes in the short-term cohort.

Variables	Univariable logistic regression analysis				Multivariable logistic regression analysis			
	OR	95% CI		*p*	OR	95% CI		*p*
		Lower	Upper			Lower	Upper	
Preop. DH	1.17	1.00	1.36	0.048	0.98	0.67	1.41	0.894
Preop. LLS	2.11	1.51	2.97	<0.001	2.01	1.23	3.28	0.005
Postop. DH	3.28	2.11	5.10	<0.001	2.81	1.40	5.65	0.004
Postop. LLS	1.78	1.00	3.15	0.050	0.87	0.24	3.25	0.855
Postop. SL	0.74	0.56	0.98	0.037	1.01	0.59	1.71	0.979
Postop. SS	0.81	0.69	0.95	0.011	0.72	0.47	1.10	0.133
Postop. PT	1.44	1.05	1.97	0.023	1.23	0.64	2.36	0.544
Insertion depth								
S group	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
L group	0.21	0.06	0.72	0.013	0.83	0.08	9.12	0.902
Facet angle	1.23	1.13	1.34	<0.001	1.20	1.02	1.41	0.030

OR, odds ratio; 95% CI, 95% confidence interval.

TABLE 5 | Univariate and multivariate logistic regression analysis of unfavorable clinical outcomes in the long-term cohort.

Variables	Univariable logistic regression analysis				Multivariable logistic regression analysis			
	OR	95% CI		*p*	OR	95% CI		*p*
		Lower	Upper			Lower	Upper	
Preop. DH	1.06	0.83	1.34	0.657	0.88	0.65	1.20	0.431
Preop. LLS	1.33	1.13	1.57	0.001	1.24	0.98	1.57	0.069
Postop. DH	1.77	1.25	2.50	0.001	1.60	1.02	2.49	0.038
Postop. LLS	2.80	1.13	6.98	0.027	1.43	0.55	3.69	0.465
Postop. SL	0.86	0.56	1.32	0.864	1.14	0.67	1.94	0.619
Postop. SS	0.88	0.68	1.12	0.289	0.91	0.67	1.24	0.566
Postop. PT	1.37	0.84	2.24	0.214	1.03	0.58	1.82	0.919
Insertion depth								
S group	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
L group	0.53	0.11	2.63	0.435	1.58	0.23	10.93	0.645
Facet angle	1.14	1.04	1.25	0.006	1.03	0.92	1.16	0.557

OR, odds ratio; 95% CI, 95% confidence interval.

the intervertebral disc (28). In the current study, outcomes in consecutive patients suggested that increased FA limits functional recovery in the short term without impacting on long-term outcomes. It was worth noting that 5–90% of chronic low back pain was triggered by facet joint pain (29). For symptomatic DLS, lumbar degenerative changes and facet joint misalignments are mutually anatomically predisposing factors (30). Thus, the facet joints, which bear approximately a quarter of the relevant axial load, play an important but neglected role in reflecting the PROs (29), and this was particularly evident in the case of short-term assessments. This variable conceals its function when focusing on and attempting to improve long-term PROs. Assuming FA does not actually worsen surgical outcomes (31), a more convincing association was acknowledged. Due to the local fixation with pedicle screws, the altered facet joint stress accelerated disc degeneration, leading to an additional influence of disc height in predicting long-term outcomes (28). This explanation was also corroborated in this study. The finite element analysis from Park et al. externally verified the homogeneity between DH and FA (32). Moreover, in the current healthcare system, pain management has become an important part of surgical intervention, especially in the treatment of degenerative spinal diseases, and has shown promising results (28, 33–35). Nevertheless, for postoperative patients, pain is a generalized clinical symptom. It is necessary for the spine surgeons to cautiously analyze the etiology of the patient's adverse PROs to achieve precise management.

In general, successful surgery is largely dependent on acquiring a solid fusion and re-establishing normal local sagittal and coronal balance. The short-term postoperative outcome was correlated with preoperative LLS but independent of postoperative LLS. In other words, the difference between the two groups postoperatively was not statistically significant (1.93 vs. 2.24, $p = 0.172$), and measurement of preoperative coronal imbalance was more valuable. The role of sagittal lumbar alignment in assessing disease grade and deciding on individualized surgical strategies was well recognized in published articles (1–3, 7), but coronal malalignment was an equally important issue to be treated with caution (1, 36, 37). Although we identified the presence of LLS preoperatively and targeted the restoration of normal lumbar coronal balance by obtaining greater lumbar lordosis intraoperatively (38), the consideration was reflected in the comparison of postoperative data, but an inferior PRO was still observed. There were several conceivable reasons for the poor outcomes. First, asymmetric degenerative changes lead to disc collapse and laxity of the paravertebral ligaments, manifested by LLS of vertebral body and lateral displacement of the disc (1). This complicated deformity had more severe pathological changes than a mere sagittal imbalance. A limited surgical benefit was observed in DLS with LLS, including an elevated and persistent back pain due to the coronal displacement of the slipped segment (36). Second, purposeful surgical strategies and instruments were recommended to correct vertebral slippage in patients with lateral slippage, including preserving the lumbar lordosis, implanting a larger cage and laterally distracting the displaced vertebral body (1, 39). However, higher restoration requirements were frequently associated with increased risk of major complications, including extensive soft tissue injury and intraoperative blood loss, all of which can reduce short-term PROs (9). Therefore, alleviating the symptoms of DLS with LLS through a modified surgical approach remained a challenge that required an all-out effort to address.

Combining SVM machine learning technique to analyze predictors of short- and long-term adverse clinical outcomes from routinely available variables was the innovation of this study. Ten-fold cross-validation AUC and confusion matrix suggested that our model performed well. However, there were several limitations in our study. First, selection bias may exist

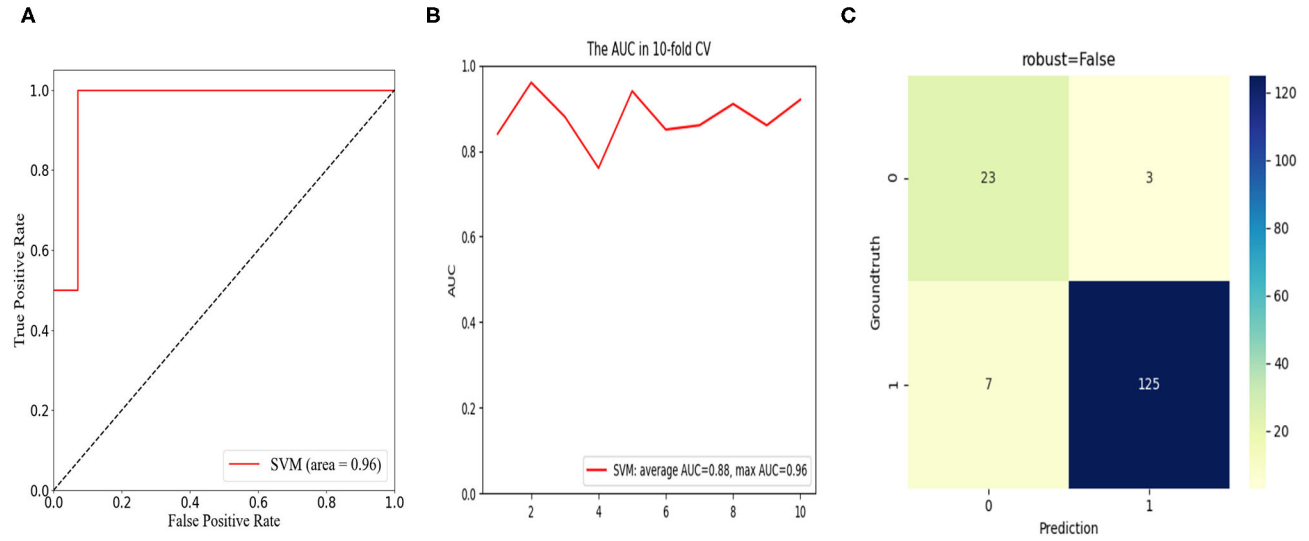


FIGURE 5 | In the short-term cohort, receiver operation characteristic (ROC) curve analysis of the SVM model with the maximum value of 0.96 **(A)**; ROC curve analysis of 10-fold cross validation of the SVM model for predicting the risk of unfavorable clinical outcomes following LIF with average AUC of 0.88 and max AUC of 0.96 **(B)**; confusion matrix of the SVM model **(C)**.

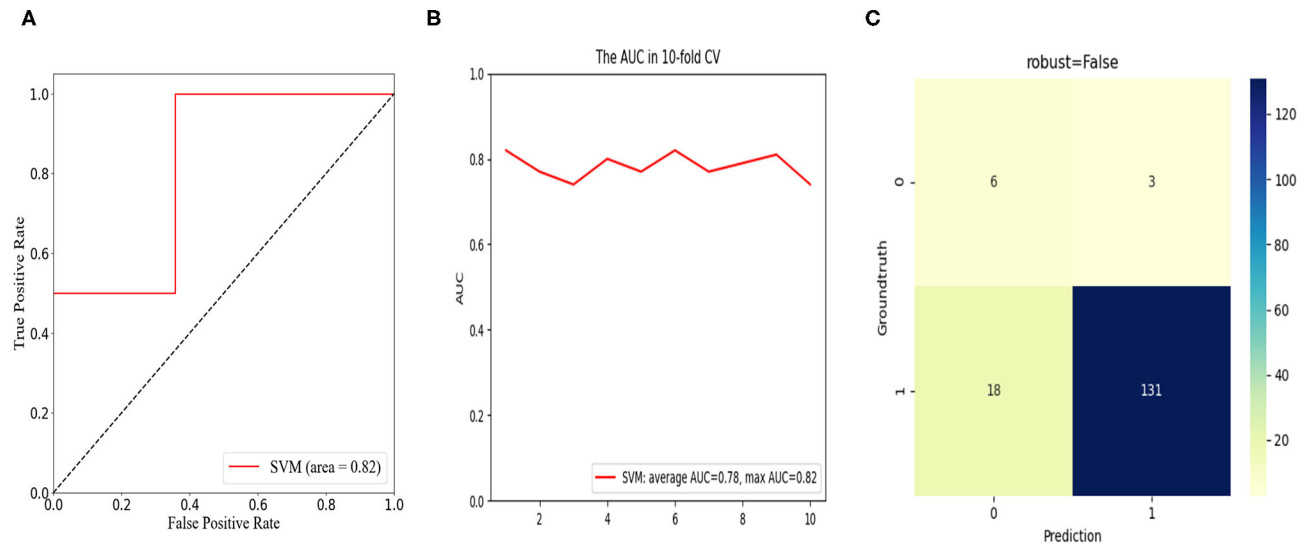


FIGURE 6 | In the long-term cohort, receiver operation characteristic (ROC) curve analysis of the SVM model with the maximum value of 0.82 **(A)**; ROC curve analysis of 10-fold cross validation of the SVM model for predicting the risk of unfavorable clinical outcomes, following LIF with average AUC of 0.78 and max AUC of 0.82 **(B)**; confusion matrix of the SVM model **(C)**.

due to the inherent flaw of retrospective studies. Second, the analyzed data were from our single institution, resulting in a lack of adequate training of the SVM model. Therefore, future studies may require a multicenter prospective randomized controlled design to assist in externally validating the credibility of our conclusion. Finally, despite an attempt to analyze extensive patient information, spinal degeneration is a multidimensional and progressive disease that still left some data unaccounted for.

CONCLUSION

In conclusion, both TLIF and PLIF are practicable surgical strategies to improve symptoms in patients with DLS. Postoperative DH, FA, and preoperative LLS were statistically associated with the short-term clinical outcome, while only postoperative DH accurately predicted the long-term clinical outcome with an average follow-up time

of 1 year. The proposed SVM model showed superior predictive performance.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was approved by an Institutional Ethics Committee at the First Affiliated Hospital of Dalian Medical University. Considering that this work was a retrospective study, the Ethics Committee waived the requirement for informed consent from patients.

AUTHOR CONTRIBUTIONS

SD and YZ completed the study design. SD, YZ, HY, and NT performed the study and collected and analyzed the data. SD, YZ, and GH drafted the manuscript. JL and KT provided the expert consultations and suggestions and conceived the study, participated in its design and coordination, and helped to embellish language. All the authors reviewed the final version of the manuscript.

FUNDING

This work is supported by National Natural Science Foundation of China (No. 81601901) and Natural Science Foundation of Liaoning (No. 2019-MS-079 and No. 20170540285).

REFERENCES

1. Pan W, Zhao JL, Xu J, Zhang M, Fang T, Yan J, et al. Lumbar alignment and patient-reported outcomes after single-level transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis with and without local coronal imbalance. *J Neurosurg Spine.* (2020) 1–7. doi: 10.3171/2020.7.SPINE20703
2. Zhong W, Liang X, Luo X, Huang T, Quan Z. Complications rate of and risk factors for the unplanned reoperation of degenerative lumbar spondylolisthesis in elderly patients: a retrospective single-Centre cohort study of 33 patients. *BMC Geriatr.* (2020) 20:301. doi: 10.1186/s12877-020-01717-2
3. Sato S, Yagi M, Machida M, Yasuda A, Konomi T, Miyake A, et al. Reoperation rate and risk factors of elective spinal surgery for degenerative spondylolisthesis: minimum 5-year follow-up. *Spine J.* (2015) 15:1536–44. doi: 10.1016/j.spinee.2015.02.009
4. Försth P, Ólafsson G, Carlsson T, Frost A, Borgström F, Fritzell P, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med.* (2016) 374:1413–23. doi: 10.1056/NEJMoa1513721
5. Ghogawala Z, Dziura J, Butler WE, Dai F, Terrin N, Magge SN, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *N Engl J Med.* (2016) 374:1424–34. doi: 10.1056/NEJMoa1508788
6. Gerling MC, Leven D, Passias PG, Lafage V, Bianco K, Lee A, et al. Risk factors for reoperation in patients treated surgically for degenerative spondylolisthesis: a subanalysis of the 8-year data from the SPORT trial. *Spine.* (2017) 42:1559–69. doi: 10.1097/BRS.0000000000002196
7. Xi X, Zeng Z, Li F, Wang C, Ma B, Xie N, et al. Caudad insertion of pedicle screws facilitates interbody distraction during spondylolisthetic vertebrae restoration: a retrospective study. *Pain Ther.* (2021) 10:1537–50. doi: 10.1007/s40122-021-00315-3
8. Zhou Q, Zhang JX, Zheng YF, Teng Y, Yang HL, Liu H, et al. Effects of different pedicle screw insertion depths on sagittal balance of lumbar degenerative spondylolisthesis, a retrospective comparative study. *BMC Musculoskelet Disord.* (2021) 22:850. doi: 10.1186/s12891-021-04736-1
9. Blumenthal C, Curran J, Benzel EC, Potter R, Magge SN, Harrington JF Jr, et al. Radiographic predictors of delayed instability following decompression without fusion for degenerative grade I lumbar spondylolisthesis. *J Neurosurg Spine.* (2013) 18:340–6. doi: 10.3171/2013.1.SPINE12537
10. Wu EQ, Hu D, Deng PY, Tang Z, Cao Y, Zhang WM, et al. Nonparametric Bayesian prior inducing deep network for automatic detection of cognitive status. *IEEE Trans Cybern.* (2021) 51:5483–96. doi: 10.1109/TCYB.2020.2977267
11. Krishnasamy L, Dhanaraj RK, Ganesh Gopal D, Reddy Gadekallu T, Aboudaif MK, Abouel Nasr E. A heuristic angular clustering framework for secured statistical data aggregation in sensor networks. *Sensors.* (2020) 20:4937. doi: 10.3390/s20174937
12. Wu EQ, Zhou M, Hu D, Zhu L, Tang Z, Qiu XY, et al. Self-Paced dynamic infinite mixture model for fatigue evaluation of pilots' brains. *IEEE Trans Cybern.* (2020). doi: 10.1109/TCYB.2020.3033005
13. Agrawal S, Sarkar S, Alazab M, Maddikunta PKR, Gadekallu TR, Pham QV. Genetic CFL: hyperparameter optimization in clustered federated learning. *Comput Intell Neurosci.* (2021) 2021:7156420. doi: 10.1155/2021/7156420
14. Ghogawala Z, Dunbar MR, Essa I. Lumbar spondylolisthesis: modern registries and the development of artificial intelligence. *J Neurosurg Spine.* (2019) 30:729–35. doi: 10.3171/2019.2.SPINE18751
15. Chan AK, Bisson EF, Bydon M, Glassman SD, Foley KT, Potts EA, et al. Laminectomy alone versus fusion for grade 1 lumbar spondylolisthesis in 426 patients from the prospective quality outcomes database. *J Neurosurg Spine.* (2018) 30:234–41. doi: 10.3171/2018.8.SPINE17913
16. Bisson EF, Mummaneni PV, Virk MS, Knightly J, Alvi MA, Goyal A, et al. Open versus minimally invasive decompression for low-grade spondylolisthesis: analysis from the quality outcomes database. *J Neurosurg Spine.* (2020) 8:1–11. doi: 10.3171/2020.3.SPINE191239
17. Bhattacharya S, Reddy Maddikunta PK, Pham QV, Gadekallu TR, Krishnan S SR, Chowdhary CL, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc.* (2021) 65:102589. doi: 10.1016/j.scs.2020.102589
18. Zahoor MI, Dou Z, Shah SBH, Khan IU, Ayub S, Reddy Gadekallu T. Pilot decontamination using asynchronous fractional pilot scheduling in massive MIMO systems. *Sensors.* (2020) 20:6213. doi: 10.3390/s20216213
19. Vogelstein JT, Bridgeford EW, Tang M, Zheng D, Douville C, Burns R, et al. Supervised dimensionality reduction for big data. *Nat Commun.* (2021) 12. doi: 10.1038/s41467-021-23102-2
20. Ayhan MS, Kühlewein L, Aliyeva G, Inhoffen W, Ziemssen F, Berens P, et al. Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection. *Med Image Anal.* (2020) 64:101724. doi: 10.1016/j.media.2020.101724
21. Yin L, Lin X, Liu J, Li N, He X, Zhang M, et al. Classification tree-based machine learning to visualize and validate a decision tool for identifying malnutrition in cancer patients. *J Parenter Enteral Nutr.* (2021) 45:1736–48. doi: 10.1002/jpen.2070
22. Duan PG, Mummaneni PV, Berven SH, Mayer R, Ruan HB, Chang CC, et al. Revision surgery for adjacent segment degeneration after fusion for lumbar spondylolisthesis: is there a correlation with roussouly type? *Spine.* (2020) 47:10–5. doi: 10.1097/BRS.0000000000003708
23. Okuda S, Nagamoto Y, Takenaka S, Ikuta M, Matsumoto T, Takahashi Y, et al. Effect of segmental lordosis on early-onset adjacent-segment disease after posterior lumbar interbody fusion. *J Neurosurg Spine.* (2021) 35:454–9. doi: 10.3171/2020.12.SPINE201888
24. Deyo RA, Martin BI, Kreuter W, Jarvik JG, Angier H, Mirza SK, et al. Revision surgery following operations for lumbar stenosis. *J Bone Joint Surg Am.* (2011) 93:1979–86. doi: 10.2106/JBJS.J.01292

25. Lombardi JS, Wiltse LL, Reynolds J, Widell EH, Spencer C 3rd. Treatment of degenerative spondylolisthesis. *Spine*. (1985) 10:821–7. doi: 10.1097/00007632-198511000-00008
26. Tian H, Wu A, Guo M, Zhang K, Chen C, Li X, et al. Adequate restoration of disc height and segmental lordosis by lumbar interbody fusion decreases adjacent segment degeneration. *World Neurosurg.* (2018) 118:856–64. doi: 10.1016/j.wneu.2018.07.075
27. Yen CP, Beckman JM, Vivas AC, Bach K, Uribe JS. Effects of intradiscal vacuum phenomenon on surgical outcome of lateral interbody fusion for degenerative lumbar disease. *J Neurosurg Spine.* (2017) 26:419–25. doi: 10.3171/2016.8.SPINE16421
28. Park JY, Kim DH, Seo DK, Yoon SH, Lee G, Lee S, et al. Predictors of response to a medial branch block: MRI analysis of the lumbar spine. *J Clin Med.* (2019) 8:538. doi: 10.3390/jcm8040538
29. Cohen SP, Raja SN. Pathogenesis, diagnosis, and treatment of lumbar zygapophysial (facet) joint pain. *Anesthesiology*. (2007) 106:591–614. doi: 10.1097/00000542-200703000-00024
30. Jentzsch T, Geiger J, Zimmermann SM, Slankamenac K, Nguyen-Kim TD, Werner CM. Lumbar facet joint arthritis is associated with more coronal orientation of the facet joints at the upper lumbar spine. *Radiol Res Pract.* (2013) 2013:693971. doi: 10.1155/2013/693971
31. Kong MH, He W, Tsai YD, Chen NF, Keorochana G, Do DH, et al. Relationship of facet tropism with degeneration and stability of functional spinal unit. *Yonsei Med J.* (2009) 50:624–9. doi: 10.3349/ymj.2009.50.5.624
32. Park WM, Kim K, Kim YH. Effects of degenerated intervertebral discs on intersegmental rotations, intradiscal pressures, and facet joint forces of the whole lumbar spine. *Comput Biol Med.* (2013) 43:1234–40. doi: 10.1016/j.compbiomed.2013.06.011
33. Zhao W, Wang Y, Wu J, Gao X, Wei Q, Lai X, et al. Long-term outcomes of epidurals with lidocaine with or without steroids for lumbar disc herniation and spinal stenosis: a meta-analysis. *Pain Physician.* (2020) 23:365–74. doi: 10.36076/ppj.2020/23/365
34. Virk S, Sandhu M, Qureshi S, Albert T, Sandhu H. How does preoperative opioid use impact postoperative health-related quality of life scores for patients undergoing lumbar microdiscectomy? *Spine J.* (2020) 20:1196–202. doi: 10.1016/j.spinee.2020.05.094
35. Peck J, Urits I, Peoples S, Foster L, Malla A, Berger AA, et al. A comprehensive review of over the counter treatment for chronic low back pain. *Pain Ther.* (2021) 10:69–80. doi: 10.1007/s40122-020-00209-w
36. Takahashi T, Hanakita J, Watanabe M, Kawaoka T, Takebe N, Kitahara T. Lumbar alignment and clinical outcome after single level asymmetrical transforaminal lumbar interbody fusion for degenerative spondylolisthesis with local coronal imbalance. *Neurol Med Chir.* (2014) 54:691–7. doi: 10.2176/nmc.st.2013-0394
37. Kato M, Namikawa T, Matsumura A, Konishi S, Nakamura H. Radiographic risk factors of reoperation following minimally invasive decompression for lumbar canal stenosis associated with degenerative scoliosis and spondylolisthesis. *Global Spine J.* (2017) 7:498–505. doi: 10.1177/2192568217699192
38. Murata Y, Takahashi K, Hanaoka E, Utsumi T, Yamagata M Moriya H. Changes in scoliotic curvature and lordotic angle during the early phase of degenerative lumbar scoliosis. *Spine.* (2002) 27:2268–73. doi: 10.1097/00007632-200210150-00016
39. Kepler CK, Rihn JA, Radcliff KE, Patel AA, Anderson DG, Vaccaro AR, et al. Restoration of lordosis and disk height after single-level transforaminal lumbar interbody fusion. *Orthop Surg.* (2012) 4:15–20. doi: 10.1111/j.1757-7861.2011.00165.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Dong, Zhu, Yang, Tang, Huang, Li and Tian. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2021.835960

Improving Alzheimer's Disease Detection for Speech Based on Feature Purification Network

Ning Liu[1,2,3], *Zhenming Yuan*[1,4]* and *Qingfeng Tang*[5]*

[1] School of Public Health, Hangzhou Normal University, Hangzhou, China, [2] Department of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, China, [3] Fujian Provincial Key Laboratory of Data-Intensive Computing, Quanzhou Normal University, Quanzhou, China, [4] School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China, [5] School of Computer and Information, Anqing Normal University, Anqing, China

Alzheimer's disease (AD) is a neurodegenerative disease involving the decline of cognitive ability with illness progresses. At present, the diagnosis of AD mainly depends on the interviews between patients and doctors, which is slow, expensive, and subjective, so it is not a better solution to recognize AD using the currently available neuropsychological examinations and clinical diagnostic criteria. A recent study has indicated the potential of language analysis for AD diagnosis. In this study, we proposed a novel feature purification network that can improve the representation learning of transformer model further. Though transformer has made great progress in generating discriminative features because of its long-distance reasoning ability, there is still room for improvement. There exist many common features that are not indicative of any specific class, and we rule out the influence of common features from traditional features extracted by transformer encoder and can get more discriminative features for classification. We apply this method to improve transformer's performance on three public dementia datasets and get improved classification results markedly. Specifically, the method on Pitt datasets gets state-of-the-art (SOTA) result.

Keywords: Alzheimer's disease, natural language processing, deep learning, transformer, machine learning, speech and language, mild cognitive impairment

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Weizheng Wang, City University of Hong Kong, Hong Kong SAR, China

***Correspondence:**
Zhenming Yuan
zmyuan@hznu.edu.cn
Qingfeng Tang
tqf1013@sina.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *15 December 2021*
Accepted: *28 December 2021*
Published: *03 March 2022*

Citation:
Liu N, Yuan Z and Tang Q (2022) Improving Alzheimer's Disease Detection for Speech Based on Feature Purification Network. Front. Public Health 9:835960. doi: 10.3389/fpubh.2021.835960

INTRODUCTION

Alzheimer's disease (AD) is a nervous degenerative disease with an insidious and irreversible onset, which is difficult to be detected in every stage. AD can influence patients' daily living ability and social communicate ability and may even lead to disability (1, 2). Researchers have found that AD has a profound impact on patients' language function (3) in addition to mood, attention, memory, movement, and so on. Language is the representation of mental activities, which can clearly reflect the relationship among language, cognition, and communication (4). Language interference is a common manifestation of patients (5) with AD which may even earlier than orientation and memory difficulties (6, 7). Picture description task, taken from Boston Aphasia Diagnostic Test (8), has already been verified sensitive to subtle cognitive deficits (9); therefore, valuable clinical information can be obtained from spontaneous speech to recognize AD. The transcripts of speech can be used to detect AD effectively.

The problem of AD recognition can be regarded as text classification problem in natural language processing (NLP). Deep learning models manifest better in classification as they can extract deep semantic features by effective model architecture automatically. For example, RNN can capture long-term dependencies within sentence, but it may neglect some important local words which may important for classification (10), and CNN can capture local and position-related features (11) but cannot give enough weight to some discriminative or special words. To solve the problem, attention mechanism was introduced. Transformer gives different weights to different words using attention mechanism, the performance of which is better than CNN and RNN. Although transformer has made great progress in producing discriminative features by powerful representation learning, there is still room to improve. There are few studies nowadays in this area to improve representation learning of deep learning. Based on GRL (12–17) in extracting common features which are not discriminative for classification, this paper proposes a novel feature purification method to improve the representation learning of transformer to get a more discriminative feature vector to diagnose AD.

The original transcripts of speech are the description of a picture, which should be comprehensive and integrated for a normal individual. That is to say, the discriminative words or sentences, with relevant and less vague words, should be included. For example, accurate descriptive words, such as "mother", "tap", "the stool is tipping", etc., are usually a better cognitive sign. Words or sentences such as "I do not know", "um", and "pause" should be an indicative of a bad cognitive condition, and they are discriminative for AD recognition. But some equivocal, inconsequential, and even irrelevant descriptions are unhelpful and may even interfere with the final classification, such as "is not that enough?", "It is great", "there may be a little breeze coming in", et al. They can disturb the representation learning of deep learning by producing suboptimal representations. To solve this problem, transformer proposes a self-attention mechanism to give weight to words and usually can get better performance than RNN and CNN. Though attention may alleviate the influence by giving a higher or lower weight for those more or less relevant words, the classification problem cannot be solved properly with inaccurate attention mechanism or specificity of data. To solve the above problems, our study, inspired by the paper (17) which used feature projection method to purify the representation learning of deep learning, proposes a novel feature purification method to improve representation learning of transformer to get more discriminative features, which is GP-Net. It has two subnetworks, a common feature learning network called G-Net and a purification network called P-Net. G-Net uses gradient reverse layer (GRL) (12, 13, 18) to extract common features which are shared by classes and have no or few roles for classification. P-Net first uses transformer encoder to extract traditional feature vector for the sentence. Then, it rules out the common features from traditional feature vector to generate more purified features. It is clear that this operation gets rid of the effect of common features and makes the system only focus on discriminative features. We will explain the principle in Method Section.

The experiments on three datasets with our method get an improved performance which prove that the purified features are more discriminative. To the best of our knowledge, there have been still no studies to recognize AD from spontaneous speech by purifying representation learning of deep learning up to now.

The key contributions we have made in this work include the following:

(1) A whole process of AD screening method, based on linguistic data, was designed and implemented.
(2) We propose a novel feature purification network to improve representation learning of transformer and get state-of-the-art (SOTA) result on Pitt dataset.
(3) The proposed method has the advantage of low cost, reliable, and convenience, which can provide a feasible solution for the screening of AD with a better performance.

RELATED WORK

Existing studies on AD diagnosis across spontaneous speech mainly focus on two aspects. One is feature extraction manually including acoustic features (19–21), linguistic features (22–25), or their combinations (21). This method is subjective and needs more professional knowledge. They are generally associated with a specific task scenario; once the scenario changes, these artificially designed features and prior settings cannot adapt to new scenarios and need to be redesigned, so the model has a low universality. The other is deep learning method which can extract deep semantic features automatically. Based on its powerful representation learning ability, the performance of deep learning is usually better than the first method. Additionally, deep learning improves the generalization ability of the classifiers which can be utilized further in different clinical environments. Deep neural network can process representation learning to extract deep semantic features using cascaded data of multilevel non-linear processing units without the need for feature engineering manually.

AD Detection Based on Deep Learning

There are many studies to detect AD from oral speech with deep learning methods (26–29), such as RNN, long–short-term memory (LSTM) networks [e.g., ELMo (30)] and CNN. Recurrent convolutional neural Networks (RCNN) (31) uses Bi-LSTM to get contextual information and then concedes the hidden output of Bi-LSTM and word embedding for classification. DPCNN (32) is a simple network with 15 layers which likes a deep CNN, and it increases network depth of CNN but does not increase the computational cost. Attention mechanism is used in many NLP tasks such as text classification (33–36). Transformer architecture [e.g., Bert (37)] uses attention mechanism to extract deep semantic features. Enhanced representation through knowledge integration (ERNIE) (38, 39), proposed in 2019 by Baidu corporation, is optimized further based on Bert model. They usually have better performance than CNN, RNN, and LSTM.

Public Dementiabank datasets or ADReSS challenge (40) datasets are often used to recognize AD. For example, Orimaye et al. (41) proposed the combination of deep language models and deep neural network to predict mild cognitive impairment (MCI) and AD. The datasets used were public Dementiabank transcript with 37 healthy elderly and 37 MCI transcripts. The study did not use any handcrafted features; just the original transcripts were fed to the model, and n-gram word embedding method combined with deep neural network (DNN) got a best AUC of 0.83. Different from our dataset and classification, there was no comparability with our method. Karlekar et al. (23) used four types of interviews: story recall, sentence construction, cookie-theft picture description, and vocabulary fluency; the dataset included 243 normal controls and 1,017 AD transcripts. Three classifiers were used for comparison, that is, LSTM-RNN, CNN, and CNN-LSTM, and achieved a best accuracy of 91.1%, but the results were somewhat questionable as mentioned in the Discussion. These methods used deep learning algorithms or their linear combinations to recognize MCI and AD. Our work is much different clearly as none of these existing studies improve representation learning of deep learning by feature purification method.

Studies Related to GRL

Our study is related to some former work. Ganin and Lempitsky (13) first introduced GRL to extract common features which were sentiment-sensitive and domain shared in domain adaptation (DA). It embeds DA to the process of representation learning in order that the final classification result is more discriminative for the domain changes. Though we use GRL to extract common features, we do not use it in the area of DA, and they also do not use for feature purification. Belinkov et al. (14) used adversarial learning to encourage the model to process representation learning on SNLI dataset. Combining with aspect attention and GRL, Kai Zhang et al. (16) studied cross domain text classification problem, and common features across domains were extracted from the aspects for text classifications. The idea of generative adversarial networks (GANs) (42) was used to ensure that the common feature space did not mix with private features and only contained pure task-independent common feature representation. In these studies, they all used GRL to extract common features inseparable for two domains, and domain-shared features were generated in the shared space according to adversarial training, whereas our study is different from them clearly as this existing work does not improve representation learning of the model. The study (17) proposed a feature projection method to further improve representation learning of deep learning from a novel angle. The method projected existing features into the orthogonal space of the common features, so the resulting projection is perpendicular to the space that common features located in and thus more discriminative for text classification. Different from this study (17) which only deletes a section of common features, we rule out the influence of whole common features, which we believe that a better classification performance should be achieved. Also, we did the experiment with the method of study (17) on Pitt dataset, and the performance is not better than our method, just as shown in **Table 1**.

TABLE 1 | Relationship between predicted and true classes.

	True class	
Predicted class	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

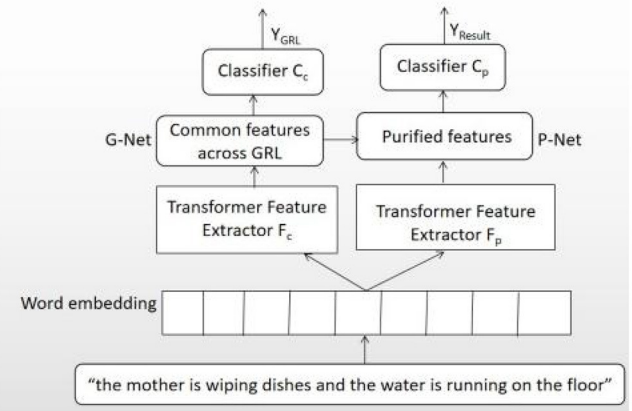


FIGURE 1 | The architecture of GP-Net.

METHODS

In this study, we propose a novel GP-Net framework to recognize AD from normal controls, which is indeed a binary classification problem in NLP.

Feature Purification Network: GP-Net

This paper proposed a novel architecture, named GP-Net, to recognize AD, the network structure of which is shown in **Figure 1**. The whole network includes two sections: G-Net and P-Net. The aim of G-Net is to extract common features by reversing the gradient direction in the training process, and these common features are shared by both classes and have no discriminative for classification. The aim of P-Net is to purify the features further by deleting the common features from traditional features extracted from transformer model. G-Net includes four sections, that is, the input layer X, feature extractor F_c, GRL, and classifier layer C_c. P-Net also includes four sections, which include the input layer X, feature extractor F_p (the features extracted by F_c and F_p have no share parameters), purification network, and classifier layer C_p. The main idea of proposed network is as follows: the feature vector, extracted by the feature extractor F_p, deletes the common features got from G-Net, and then, more discriminative purified features have got for the final classification. Two operations, including G-Net and P-Net, are required in order for feature purification operation.

Transformer Extractor

This study uses transformer encoder as the feature extractor. Transformer is a SOTA model which has a novel architecture to solve sequence to sequence tasks. The model can capture long-distance dependencies and learn global semantic features of input text thoroughly through multihead self-attention mechanism. As transformer has some mechanisms as self-attention and location code, it has excellent feature extraction and semantic abstract competence. Like most Seq2Seq model, transformer model also uses encoder–decoder structure, the encoder of which is a better feature extractor with multihead attention and feed forward neural network.

Supposing G-Net and P-Net have the same input X, the feature extractors of G-Net and P-Net are F_c and F_p, which can get the advanced features f_p and f_c from the input layer, respectively, but there are not any shared parameters between them. We refer to the features of P-Net and G-Net, respectively, as

$$f_p = Transformer_p(X), \tag{1}$$
$$f_c = Transformer_c(X), \tag{2}$$

Additional details of G-Net and P-Net will be introduced in G-Net and P-Net module.

G-Net Module

The main goal of G-Net module is to extract common features among datasets, which is not discriminative for the classification. As common features are those shared by all the classes, the classifier cannot use them to distinguish different classes effectively. To get common features, GRL (12, 13, 18) is added between the feature extractor F_c and the classifier to reverse the gradient direction. The common features that are shared among different classes are obtained after the training module. G_λ can be thought as two incompatible equations that describe the forward and back propagation behaviors:

$$G_\lambda(x) = x, \tag{3}$$
$$\frac{\partial G_\lambda}{\partial x} = -\lambda I, \tag{4}$$

where λ is a hyper parameter. We process feature vector f_c through GRL and get f_c', for example, G_λ (f_c) = f_c'. To make f_c' close to real common features, GRL acts as identity transform during the forward propagation and then takes the gradient from subsequent level and changes the value (i.e., multiplies it by -λ) before passing it to the next layer during back propagation, and this operation can ensure that the feature distributions are similar and as indistinguishable as possible for the classifier. Only in this way we can get the common features sharing among classes. Finally, f_c' is fed to classifier C_c.

$$Y_{GRL} = softmax(W_c^* f_c' + b_c), \tag{5}$$
$$Loss_c = CrossEntropy(Y_{True}, Y_{GRL}), \tag{6}$$

where W_c and b_c are the weight and bias of classifier C_c. By optimizing $Loss_c$, the feature extractor F_c can extract common features of different classes.

P-Net Model

The goal of P-Net is to extract the semantic information from input example first and then purify features for the classification. Supposing the traditional feature vector we extracted by transformer is f_p, the common feature vector is f_c. The final feature vector for classification is f_w.

$$f_w = f_p - f_c, \tag{7}$$

As f_c disturbs the classification result, we delete f_c from f_p to eliminate the influence of nondiscriminative feature vector (i.e., common features), so the feature vector f_w is more discriminative than f_p. Finally, the purification feature vector f_w is fed to classifier C_p.

$$Y_{Result} = soft\ max(W_p * f_w + b_p), \tag{8}$$
$$Loss_p = CrossEntropy(Y_{True}, Y_{Result}), \tag{9}$$

where W_p and b_p are the weight and bias of classifier C_p. By optimizing $Loss_p$, the feature extractor F_p can purify the features, $Loss_c$ and $Loss_p$ are trained simultaneously, but they use different optimizers. $Loss_c$ use moment SGD optimizer because Ganin and Lempitsky (13) also used moment SGD, and $Loss_p$ use Adam optimizer. We also conducted the experiments using Adam optimizer for both G-Net and P-Net and found that the results made no difference when using two different optimizers. In terms of optimization targets of feature extractor F_c, though the two losses are opposite to each other, a balance can be found to make the extracted feature f_c closer to real common features. The algorithm description of the whole training process is shown in **Algorithm 1**:

Algorithm 1 GP-Net

1: Input:
Supposing the datasets are $D = \{(x_i, y_i)\}_{i=1}^N$, x_i is the embedding matrix of deep learning, $X_i \in R^{Lk}$, y_i is the corresponding classes; randomly initialized the parameters of GP-Net.
2: **For** every iteration b=1, 2,...N, **do**:
3: Sample one batch x_b from D,
4: **G-Net section:**
5: Generate common features (CFs) (Equation 1)
6: CFs go through GRL and get the features closer to the common features (Equation 3)
7: Do the classification (Equation 5)
8: **P-Net section:**
9: Generate traditional features(tfs) (Equation 2)
10: Get purified features (Equation 7)
11: Do the classification (Equation 8)
12: **Update parameters:**
13: The parameters of P-Net and G-Net are updated together (Equation 6 and Equation 9)
14: **End for**

TABLE 2 | AD vs. CTRL classification scores on Pitt datasets.

Method	Embedding	Classifier	Precision	Recall	Accuracy	F1
Sweta Karlekar (23)	POS	CNN-RNN	-	-	91.1	-
Fritsch et al. (29)	n-gram	NNLM+LSTM	-	-	85.6	-
Orimaye et al. (41)	n-grams	D2NN	-	-	88.9	-
Fraser et al. (43)	35 Hand-Crafted Feature	LR	-	-	81.92	-
Yancheva et al. (45)	12 Cluster-Based Features+LS&A	Random Forest	80.00	80.00	80.00	80.00
Sirts et al. (46)	Cluster+PID+SID Features	LR	74.4 ± 1.5	72.5 ± 1.2	-	72.7 ± 1.2
Hernandez et al. (47)	105 Hand-Crafted Features	SVM	81.00	81.00	79.00	81.00
Roshanzamir et al. (48)	BERT $_{Base}$ (Sentence Level)	LR	90.31 ± 7.36	76.52 ± 8.06	84.46 ± 6.31	82.72 ± 7.21
Roshanzamir et al. (48)	Bert $_{Large}$	LR	90.57 ± 3.18	84.34 ± 7.58	88.08 ± 4.48	87.23 ± 5.20
Pan et al. (49)	GloVe Word Embedding Sequence	BiLSTM|GRU Hierarchical Attention	84.02	84.97	-	84.43
Li et al. (50)	185Hand-Craft Features	LR	-	-	77	-
Fraser et al. (51)	Info and LM Features	SVM	-	-	75	77
Transformer+FP[25]	Transformer +Feature projection	Transformer	88	**91**	90.3	90.6
Transformer+GP	Transformer+Feature purification	Transformer	**94**	89	**93.5**	**91.19**

The study marked with bold is the best performances on Pitt dataset.

EXPERIMENT

Datasets

Three datasets are used to carry out the experiments which are all the dialogues of picture description task, including English and Chinese.

Pitt Datasets

This is a Pitt corpus (43) from Dementiabank dataset (43), which comes from a study at School of Medicine in Pittsburgh University and is gathered longitudinally every year. More detailed description about the dataset can refer to the study (43). After deleting some unqualified datasets such as unknown label, memory impairment, and other dementia diagnose, for example, vascular dementia, there are 498 participants enrolled in our study after data preprocessing, which is composed of 242 controls and 256 possible or probable AD. Both categories are balanced basically.

ADReSS Datasets

The datasets include 78 dementia patients and 78 normal controls from ADReSS challenge in 2020. The speech is segmented using a voice activity detect method based on signal energy value. All datasets have already been preprocessed by removing noise.

iFLY Datasets

The Chinese datasets include 111 CTRL and 68 AD, with 60 women and 51 men in CTRL group and 38 women and 30 men in AD group, respectively. More details can refer to the website: http://challenge.xfyun.cn/2019/gamedetail?blockId=978.

Feature Parameters

In the training stage of GP-Net module, a stochastic gradient of 0.9 is used as momentum, and annealing learning rate can be calculated by the following formula:

$$l_p = \frac{l_0}{(1 + \alpha^* p)^\beta} \tag{10}$$

where $l_0 = 0.01$, $\alpha = 10$, $\beta = 0.75$, p is training progress linearly changing from 0 to 1. In Equation (4), the parameter λ is set as [0.05, 0.1, 0.2, 0.4, 0.8, and 1.0]. Transformer encoder is used as the feature extractor, with three blocks and single head specifically.

Experiment Results

For our model, 5-fold cross validation was used for training dataset. The dataset was divided into five parts randomly, of which four parts were used for training, and one part was used for test. We repeat the process five times using different test dataset

TABLE 3 | The result of pre-trained models on Pitt dataset.

Model	Embedding	Classifier	Precision	Recall	Accuracy	F1
BertCNN	Bert	CNN	58.85	56.25	56.25	52.79
BertRCNN	Bert	RCNN	-	50.00	50.00	33.33
BertDPCNN	Bert	DPCNN	41.11	47.92	47.92	35.59
ERNIEDPCNN	ERNIE	DPCNN	-	50.00	50.00	33.33
BertLogistic	Bert	Logistic Regression	88	85	86.20	85.60
Transformer+GP	Transformer	Transformer	94.00	89.00	93.50	91.19

every time. Finally, the results of five times were summarized, and the average value was used as the estimation of model performance index. The classification of our model adopts the following indexes: accuracy, precision, recall, and F1 score are used as the final index (44). The relationship between the actual class and predicted class is shown in **Table 1**, and the evaluate metrics in this study are defined as Equations (11–14).

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{Recall Rate} = \frac{TP}{TP + FN} \tag{13}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{14}$$

Table 2 is the classification scores for AD and CTRL on Pitt Dementiabank datasets, including handcrafted features extracted methods and deep learning methods. As far as we know, SOTA on Pitt corpus is the study of Roshanzamir et al. (48) in 2021, and our method in this paper performs better than SOTA. Also, transformer+FP[25] is the feature project method for text classification, and we did the experiment with this method on our Pitt datasets; the performance of our method is better than the project method with the same datasets. To further compare with some proposed popular pretrained models in recent years, including Bert (37), ERNIE (38), RCNN (31), and DPCNN (32), we do the comparative experiments with the combination models, including BertRCNN, BertDPCNN, BertLogistic, and ERNIEDPCNN models, which are the combination of Bert + CNN, Bert + RCNN, Bert + DPCNN, Bert + Logistic Regression, and ERNIE + DPCNN, respectively. The former is the feature extractor and the latter is the classifier. The evaluation index is shown in **Table 3**.

From **Table 3**, we can find that the performance of first four models is not better, with an accuracy of only 50% or so, BertLogistic model has a better accuracy of 86.2%, and our method gets the best result than these pretrained models. In the meanwhile, to prove the superior of our method, we also test the method on ADReSS and iFLY datasets. The result is shown in **Table 4**, accuracy is improved by 2.1, 4.3, and 2.1%, respectively, on Pitt, ADReSS, and iFLY dataset, which means that the purified features are more discriminative than the features extracted by transformer encoder. Though we do not get SOTA accuracy on ADReSS and iFLY dataset, the performance of our method is improved than transformer.

TABLE 4 | Accuracy on three datasets.

Model	Pitt	ADReSS	iFLY
Transformer	91.4	74.3	81.6
Transformer+GP	93.5	78.6	83.7

DISCUSSION

Why the performance is better after purification? We know that transformer is superior to RNN, CNN on its long-distance reasoning ability, but it is not easy to understand the deep semantic feature vector extracted by transformer as deep learning is a "black-box." The common features in the study are the vector that cannot differentiate for classification in semantic space. They may be the words or sentences that are unimportant, unmeaningful, and irrelevant that may disturb the final classification. Our original dataset is a dialogue of description. It should include some important people, scenes, and ongoing events in the picture. The study (52) pointed out the seed words of the picture should include the following 23 words: boy, girl, woman, cookie, stool, sink, overflow, fall, window, curtain, plate, cloth, jar, water, cupboard, dish, kitchen, garden, take, wash, reach, attention, and see. The sentences including these words are helpful for the classification. Other unrelated words or sentences such as "Can you tell me", "look, there is no people outside" which are unhelpful words or sentences that cannot distinguish cognitive condition. They are, which we think maybe the common features, unhelpful and may even disturb the final classification. When we rule out these words or sentences (i.e., common features) that disturb the classification, the result can be improved correspondingly. The features extracted manually in this area usually include part-of-speech (POS), fluency, semantic feature, lexical richness, and so on. Now, there is an opinion that the features that deep learning extracted automatically maybe are much like the features that people extract manually, and deleting those unhelpful words for the classification can improve the classification performance.

We know that in transformer model, complexity per layer of self-attention is $O(n^2 * d)$, where d is the representation dimension, and n is the sequence length. Our model includes

two sections, one is transformer encoder, the other is feature purification layer which just multiply (-λ) when running back propagation. Both of them can run concurrently and have the same complexity, so the computational complexity of our model is the same as that of self-attention.

CONCLUSIONS

Nowadays, many medical problems used artificial intelligence method to solve (53, 54). which is low cost and convenient. Two methods, that is, feature extraction manually and automatically by deep learning, are usually used to recognize disease. Features extracted method manually based on machine learning does not generalize well, as it needs many special knowledge and annotation to extract features. Due to high cost of manual annotation, it is not feasible to procure numbers of annotated datasets for most clinical tasks. But deep learning does not need any annotation and can finish the process automatically. This paper combines transformer-based model with a feature purification network to improve the classification performance to a large extent. We pretrain transformer and then fine-tune the model on new datasets to transfer learned knowledge to our text classification task. Our work is obviously different from the former studies in AD recognition because none of the former studies improve representation learning of deep learning in this area, as far as we know. The common features extracted by GRL maybe the words that shared by different classifications, or nonimportant words that have small role for classification, ruling out them from traditional representation vector can improve the performance of the model. In addition, we can develop WeChat procedure or APP in mobile device further in order that the elderly can test their cognitive condition at home. So, large volumes of patient's datasets need to be transferred to central cloud server for data analysis, the safety of which is important, and blockchain technology is a better choice which may ensure the security of medical data (53, 55).

Transformer model is still the most widely used deep learning algorithm, but the time complexity of self-attention is higher, which hinders the development of the model, so the improvement of model efficiency is of great importance in the future. Transformer, as the feature extractor we used in this study, can also be replaced by other deep learning algorithms such as Bert, RNN, CNN, and so on; next, we will perfect the work further. In the meanwhile, we also believe that our feature purification method may predict other diseases that language and cognitive impairment related, such as Parkinson's disease, Aphasia, and Autism spectrum disorder. Aphasia is maybe more pronounced as Aphasia is a disease of the brain tissue associated with language function. Our method provides a feasible solution for detecting patients with AD at the doorsteps. Feature purification method for deep learning, as far as we think, is a promising direction to explore in the future.

DATA AVAILABILITY STATEMENT

The public datasets we used can get from the website: https://sla.talkbank.org/TBB/dementia/English/Pitt, or visit our github website: https://github.com/lzy1012/Public-Pitt-Dementiabank-Dataset.

AUTHOR CONTRIBUTIONS

ZY designed the research. QT analyzed the data and interpreted the analysis. NL and ZY wrote the main manuscript text and revised carefully. All authors reviewed and approved the final manuscript.

FUNDING

This research was funded by grants of Natural Science Foundation of Zhejiang Province (LGF20F020009), Anhui Provincial Natural Science Foundation (No. 2108085QF269), and the 4th Graduate Student Innovation and Entrepreneurship Competition of Hangzhou Normal University.

REFERENCES

1. Sousa RM, Ferri CP, Acosta D, Albanese E, Guerra M, Huang YQ, et al. Contribution of chronic diseases to disability in elderly people in countries with low and middle incomes: a 10/66 dementia research group population-based survey. *Lancet*. (2009) 374:1821–30. doi: 10.1016/S0140-6736(09)61829-8
2. Zhou Y, Lu Y, Pei Z. Intelligent diagnosis of Alzheimer's disease based on internet of things monitoring system and deep learning classification method. *Microprocess Microsyst*. (2021) 83:104007. doi: 10.1016/j.micpro.2021.104007
3. Sabat SR. Language function in Alzheimer's disease: a critical review of selected literature. *Lang Commun*. (1994) 14:331–51. doi: 10.1016/0271-5309(94)90025-6
4. Bartha L, Benke H. Acute conduction aphasia: an analysis of 20 cases. *Brain Lang*. (2003) 85:93–108. doi: 10.1016/S0093-934X(02)00502-3
5. Appell J, Kertesz A, Fisman M. A study of language functioning in Alzheimer patients. *Brain Lang*. (1982) 17:73–91. doi: 10.1016/0093-934X(82)90006-2
6. Corkin S DK, Growdon J, Usdin E, Wurtman R. *Some Relationships Between Global Amnesias and the Memory Impairments in Alzheimer's Disease*. Raven Press (1982). p. 149–64.
7. Wechsler AF, Verity MA, Rosenschein S, Fried I, Scheibel AB. Pick's disease. A clinical, computed tomographic, and histologic study with Golgi impregnation observations. *JAMA Neurology*. (1982) 39:287–90. doi: 10.1001/archneur.1982.00510170029008
8. Goodglass H, Kaplan E, Barresi B. *Boston Diagnostic Aphasia Examination Record Booklet*. Lipppincott Williams and Wilkins (2001).
9. Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *J Clin Exp Neuropsychol*. (2008) 30:501–56. doi: 10.1080/13803390701550128
10. Yin W, Kann K, Yu M, Hinrich S. *Comparative Study of CNN and RNN for Natural Language Processing. arXiv [Preprint] arXiv*:1702.01923 (2017).
11. Scherer D, Muller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. *ICANN*. (2010) 19(Suppl 8):92–101. doi: 10.1007/978-3-642-15825-4_10

12. Ganin Y, Lempitsky V. Unsupervised domain adaptation by back propagation. In: *International Conference on Machine Learning*. Florida, FL: PMLR (2015). p. 1180–9.
13. Liu P, Qiu X, Huang X. Adversarial multi-task learning for text classification. *arXiv [Preprint]*. (2017). arXiv:1704.05742. doi: 10.18653/v1/p17-1001
14. Belinkov Y, Poliak A, Shieber SM, Durme BV, Alexander MR. On adversarial removal of hypothesis-only bias in natural language inference. *arXiv [Preprint] arXiv*:1907.04389 (2019). doi: 10.18653/v1/S19-1028
15. Qin L, Xu X, Che W, Zhang Y, Liu T. Dynamic fusion network for multi-domain end-to-end task-oriented dialog. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020). doi: 10.18653/v1/2020.acl-main.565
16. Zhang K, Zhang H, Liu Q, Zhu HS, Chen E. Interactive attention transfer network for cross-domain sentiment classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, HI (2019). doi: 10.1609/aaai.v33i01.33015773
17. Qin Q, Hu W, Liu B. Feature projection for improved text classification. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020). doi: 10.18653/v1/2020.acl-main.726
18. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F. Domain-adversarial training of neural networks. *J Mach Learn Res*. (2016) 17:2096–30. doi: 10.1007/978-3-319-58347-1_10
19. Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Front Aging Neurosci*. (2018) 10:369. doi: 10.3389/fnagi.2018.00369
20. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis*. (2015) 49:407–22. doi: 10.3233/JAD-150520
21. Gosztolya G, Vincze V, Tóth L, Pákáski M, Hoffmann I. Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Comput Speech Lang*. (2018) 53:181–97. doi: 10.1016/j.csl.2018.07.007
22. Fraser KC, Fors KL, Kokkinakis D. Multilingual word embeddings for the assessment of narrative speech in mild cognitive impairment. *Comput Speech Lang*. (2018) 53:121–39. doi: 10.1016/j.csl.2018.07.005
23. Karlekar S, Niu T, Bansal M. Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. New Orleans, Louisiana (2018). p. 701–7.
24. López-De-Ipiña K, Alonso JB, Solé-Casals J, Barroso N, Henriquez P, Faundez-Zanuy M. On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. *Cognitive Computation*. (2015) 7:44–55. doi: 10.1007/s12559-013-9229-9
25. Orimaye SO, Wong SM, Wong CP. Correction: deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia. *Plos ONE*. (2019) 14:e0214103. doi: 10.1371/journal.pone.0214103
26. Fritsch J, Wankerl S, Nöth E. Automatic diagnosis of Alzheimer's disease using neural network language models. In: *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton). (2019).
27. López-de-Ipiña K, Martinez-de-Lizarduy U, Calvo PM, Beitia B, Faundez ZM. Analysis of disfluencies for automatic detection of Mild Cognitive Impartment: a deep learning approach, 2017. In: *International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. (2017). p. 1–4. doi: 10.1109/IWOBI.2017.7985526
28. Palo FD, Parde N. Enriching neural models with targeted features for dementia detection. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. Florence (2019). doi: 10.18653/v1/P19-2042
29. Fritsch J, Wankerl S, Noth E. Automatic diagnosis of Alzheimer's disease using neural network language models. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brighton). (2019). p. 5841–5.
30. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K. Deep contextualized word representations. *arXiv [Preprint] arXiv*:1802.05365 (2018).
31. Lai S, Xu L, Liu K, Zhao J. Recurrent convolutional neural networks for text classification. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Austin, TX (2015). p. 2267–73
32. Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. (2017). p. 562–70. doi: 10.18653/v1/P17-1052
33. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv [Preprint]*. (2016). arXiv: 1409.0473v7. Available online at: https://arxiv.org/pdf/1409.0473.pdf (accessed May 19, 2016).
34. Yang Z, Yang D, Dyer C, He XD, Smola A, Hovy E. Hierarchical attention networks for document classification. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. (2016). p. 1480–9. doi: 10.18653/v1/N16-1174
35. Lin Z, Feng M, Santos CN, Yu M, Xiang B, Zhou B. A structured self-attentive sentence embedding. *arXiv [Preprint] arXiv*:1703.03130 (2017).
36. Ma Y, Peng H, Cambria E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: *Thirtysecond AAAI conference on artificial intelligence*. New Orleans, LA (2018).
37. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv [Preprint] arXiv*:1810.04805 (2018).
38. Zhang Z, Han X, Liu Z, Jiang X, Sun M, Liu Q. ERNIE: enhanced language representation with informative entities. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence. (2019). p. 1441–51.
39. Liu N, Yuan ZM. Spontaneous language analysis in Alzheimer's disease: evaluation of natural language processing technique for analyzing lexical performance. *J Shanghai Jiaotong Univ*. (2021) 1–8. doi: 10.1007/s12204-021-2384-3
40. Luz S, Haider F, Fuente S, Fromm D, Macwhinney B. *Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge*. Shanghai, China: Interspeech.2020-2571 (2020). doi: 10.21437/Interspeech.2020-2571
41. Orimaye SO, Wong SM, Wong CP, Liang P. Deep language space neural network for classifying mild cognitive impairment and Alzheimer-type dementia. *PLoS ONE*. (2018) 13. doi: 10.1371/journal.pone.0205636
42. Zhao W, Gao H, Chen S, Wang N. Generative Multi-Task Learning for Text Classification. *IEEE Access*. (2020) 8:86380–7. doi: 10.1109/ACCESS.2020.2991337
43. Becker JT, Boiler F, Lopez OL, Saxton J, Mcgonigle KL. The Natural History of Alzheimer's Disease: Description of Study Cohort and Accuracy of Diagnosis. *Arch Neurol*. (1994) 51:585–94. doi: 10.1001/archneur.1994.00540180063015
44. Na KS, Cho SE, Zong WG, KimYK. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. *Neurosci Lett*. (2020) 721:134804. doi: 10.1016/j.neulet.2020.134804
45. Yancheva M, Rudzicz F. Vector-space topic models for detecting Alzheimer's disease. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. (2016). p. 2337–46. doi: 10.18653/v1/P16-1221
46. Sirts K, Piguet O, Johnson M. Idea density for predicting Alzheimer's disease from transcribed speech. *arXiv [Preprint]*. (2017). arXiv:1706.04473v1. doi: 10.18653/v1/K17-1033
47. Hernández-Domínguez L, Ratté S, Sierra-Martínez G, Roche-Berguac A. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *DADM*. (2018) 10:260–8. doi: 10.1016/j.dadm.2018.02.004
48. Roshanzamir A, Aghajan H, Baghshah MS. Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Medical Inform Decis Mak*. (2021) 21:1–14. doi: 10.1186/s12911-021-01456-3
49. Pan Y, Mirheidari B, Reuber M, Venneri A, Blackburn D, Christensen H. Automatic hierarchical attention neural network for detecting AD. *Proc Interspeech*. (2019) 2019:4105–9. doi: 10.21437/Interspeech.2019-1799
50. Li B, Hsu YT, Rudzicz F. Detecting dementia in mandarin Chinese using transfer learning from a parallel corpus. *arXiv [Preprint] arXiv:1903.00933*. (2019). doi: 10.18653/v1/N19-1199

51. Fraser KC, Linz N, Li B, Fors KL, Rudzicz F, König A, et al. Multilingual prediction of Alzheimer's disease through domain adaptation and conceptbased language modelling. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Florence. (2019). p. 3659–70.
52. Bucks RS, Singh S, Cuerden JM, Wilcock GK. Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology*. (2000) 14:71–91. doi: 10.1080/026870300401603
53. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Lwendi C, Wei C. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) 16:647–58. doi: 10.1049/ipr2.12222
54. Thippa RG, Bhattacharya S, Maddikunta P, Hakak S, Tariq U. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimed Tools Appl*. (2020) 26:1–25. doi: 10.1007/s11042-020-09988-y
55. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet Things J*. (2021). doi: 10.1109/JIOT.2021.3117762

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Yuan and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2021.833970

A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Clear Cell Renal Cell Carcinoma: A Population-Based Study

*Chenghao Zhanghuang[1,2], Jinkui Wang[1], Zhaoxia Zhang[1], Liming Jin[1], Xiaojun Tan[1,3], Tao Mi[1], Jiayan Liu[1], Mujie Li[1] and Dawei He[1]**

[1] Department of Urology, Chongqing Key Laboratory of Children Urogenital Development and Tissue Engineering, Chongqing Key Laboratory of Pediatrics, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China, [2] Yunnan Key Laboratory of Children's Major Disease Research, Department of Urology, Kunming Children's Hospital (Children's Hospital Affiliated to Kunming Medical University), Kunming, China, [3] Department of Urology, Nanchong Central Hospital, The Second Clinical Medical College, North Sichuan Medical University, Nanchong, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Abdul Rehman Javed,
Air University, Pakistan
Sharnil Pandya,
Symbiosis International University, India

***Correspondence:**
Dawei He
hedawei@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 12 December 2021
Accepted: 28 December 2021
Published: 03 March 2022

Citation:
Zhanghuang C, Wang J, Zhang Z, Jin L, Tan X, Mi T, Liu J, Li M and He D (2022) A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients With Clear Cell Renal Cell Carcinoma: A Population-Based Study. Front. Public Health 9:833970. doi: 10.3389/fpubh.2021.833970

Background: Clear cell renal cell carcinoma (ccRCC) is expected in the elderly and poor prognosis. We aim to explore prognostic factors of ccRCC in the elderly and construct a nomogram to predict cancer-specific survival (CSS) in elderly patients with ccRCC.

Methods: Clinicopathological information for all elderly patients with ccRCC from 2004 to 2018 was downloaded from the Surveillance, Epidemiology, and End Results (SEER) program. All patients were randomly assigned to a training cohort (70%) or a validation cohort (30%). Univariate and multivariate Cox regression models were used to identify the independent risk factors for CSS. A new nomogram was constructed to predict CSS at 1-, 3-, and 5 years in elderly patients with ccRCC based on independent risk factors. Subsequently, we used the consistency index (C-index), calibration curves, and the area under the receiver operating curve (AUC) and decision curve analysis (DCA) to test the prediction accuracy of the model.

Results: A total of 33,509 elderly patients with ccRCC were enrolled. Univariate and multivariate Cox regression analyses results showed that age, sex, race, marriage, tumor size, histological grade, tumor, nodes, and metastases (TNM) stage, and surgery were independent risk factors for CSS in elderly patients with ccRCC. We constructed a nomogram to predict CSS in elderly patients with ccRCC. The C-index of the training cohort and validation cohort was 0.81 (95% CI: 0.802–0.818) and 0.818 (95% CI: 0.806–0.830), respectively. The AUC of the training cohort and validation cohort also suggested that the prediction model had good accuracy. The calibration curve showed that the observed value of the prediction model was highly consistent with the predicted value. DCA showed good clinical application value of the nomogram.

Conclusion: In this study, we explored prognostic factors in elderly patients with ccRCC. We found that age, sex, marriage, TNM stage, surgery, and tumor size were independent

risk factors for CSS. We constructed a new nomogram to predict CSS in elderly patients with ccRCC with good accuracy and reliability, providing clinical guidance for patients and physicians.

Keywords: nomogram, clear cell renal cell carcinoma, cancer-specific survival, elderly patients, SEER

INTRODUCTION

Renal cell carcinoma (RCC) is a malignant tumor originating from renal tubular epithelial cells. More than 400,000 cases are diagnosed each year, and RCC incidence is about twice as high in men as in women (1, 2). There are several major pathological subtypes of RCC, of which clear cell renal cell carcinoma (ccRCC) accounts for 70–80%, with papillary renal cell carcinoma (PRCC) and chromophobe renal cell carcinoma (CRCC) remaining (3, 4). ccRCC is characterized by rich glycogen and lipids in cells and is manifested as chromosomal 3p deletion and genomic mutation in Von Hippel-Lindau (VHL) tumor suppressor allele (5). Although ccRCC is a disease that can be cured by early surgery, such as radical nephrectomy (RN), partial nephrectomy (PN), and local tumor excision represented by radiofrequency ablation (RFA), recurrence and metastasis occur in up to one-third of cases (6). Moreover, these metastases often suggest poor prognosis, which makes the prognosis of ccRCC significantly different from that of other RCCs (7).

Age is a significant risk factor for cancer. With the increase of age, the probability of cell mutation caused by molecular changes also increases, and cell mutation is generally considered the initiation event of cancer (8). The elderly over 65 years old account for more than 75% of the diagnosed patients with ccRCC. With the aging population and the extension of life expectancy, the incidence rate of renal cancer in the elderly is still increasing year by year (2).

Unfortunately, until today, studies on genetics (9), molecular structure (10), and clinicopathological features (11) have not been able to predict the clinical outcome and prognosis of ccRCC patients accurately. Accurate prediction is crucial in clinical decision-making, patient confidence building, and medical care improvement. In recent years, the application of nomogram prediction models, such as UISS (12), the Stage, Size, Grade, and Necrosis (SSIGN) (13), and Leibovich (14), in the prognosis of cancer patients provides clinicians with a basis for precise and individualized treatment of various cancer patients.

At present, artificial intelligence has been widely used in human health. Awais et al. (15) used texture analysis techniques in fluorescence imaging to help dentists identify areas of oncogenic cavity abnormalities in the oral cavity, and thus, perform biopsies more efficiently to predict the histological diagnosis of epithelial dysplasia. Shah et al. (16) used machine learning to detect the condition and risk of cardiac arrest early to improve survival in patients with heart disease.

However, to our knowledge, no predictive model for cancer-specific survival (CSS) of elderly ccRCC has been reported. Moreover, the prognostic factors of survival in elderly patients with ccRCC remain unclear. We used big data from the Surveillance, Epidemiology, and End Results (SEER) program for the first time to explore factors influencing CSS in elderly patients with ccRCC. Based on the above situation, we developed the nomogram prediction model. We validated its accuracy in evaluating the CSS rate of patients with ccRCC, providing a reference for clinical diagnosis and treatment.

PATIENTS AND METHODS

Data Source and Data Extraction

We downloaded clinicopathological information for all elderly patients with ccRCC from 2004 to 2018 from the SEER program of the National Cancer Institute. SEER database is the National cancer database of the United States, covering about 30% of the population and containing 18 cancer registries. Demographic information, clinicopathological information, and follow-up data for all cancer patients are publicly available from the SEER database. Because patient data are publicly available and cannot be identified, our study needs not require ethical approval or informed consent from patients. Our research methods comply with the regulations of the SEER database.

All the demographic information of ccRCC in elderly patients (age, sex, race, year of diagnosis, and marriage), clinical pathologic information (laterality, tumor size, histological grade, and tumor, nodes, and metastases [TNM] stage), treatment (surgery, radiation therapy, and chemotherapy), follow-up information (survival status, survival time, and the cause of death) were collected. Inclusion criteria were as follows: (1) pathological diagnosis of ccRCC (International Classification of Diseases for Oncology [ICD-O]-3 codes, 8310); (2) age $\geq$65; (3) the years of diagnosis were 2004–2018; and (4) unilateral renal tumor. Exclusion criteria were as follows:(1) tumor histological grade is unknown; (2) tumor size is unknown; (3) unknown surgical method; (4) survival time <1 month; and (5) incomplete follow-up information. The patient screening flow chart is shown in **Figure 1**.

The patients' race was divided into three categories: white, black, and others (American Indian/AK Native, Asian/Pacific Islander). Patients were divided into years of diagnosis from 2004 to 2010 and 2011 to 2018. The patient's tumor histological grades included grade I (well-differentiated), grade II (moderately differentiated), grade III (poorly differentiated), and grade IV (undifferentiated). Patients were classified as non-operative (SEER Operation Code 0), local tumor excision (SEER Operation Code 10-27), PN (SEER Operation Code 30), and RN (SEER Operation Code 40-80).

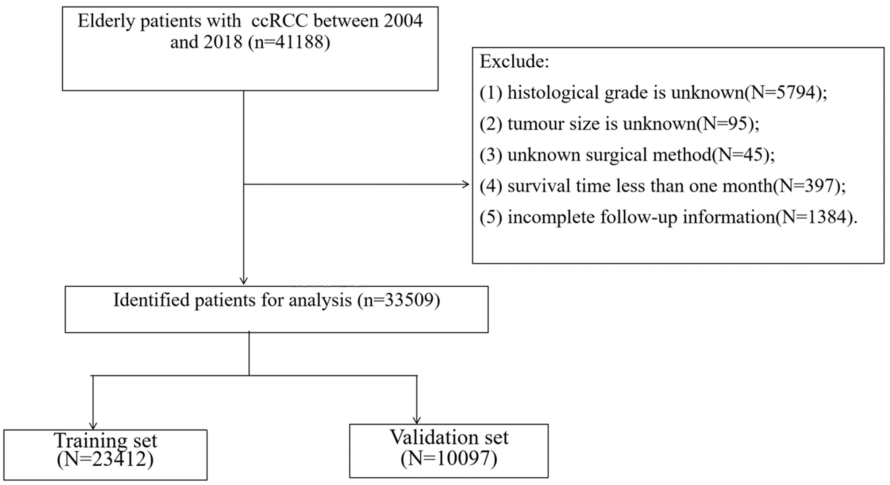


FIGURE 1 | Flowchart for inclusion and exclusion of all patients.

Univariate and Multivariate Cox Regression Analyses

All elderly patients with ccRCC were randomly assigned to a training cohort (70%) or a validation cohort (30%). All variables that might be associated with patient outcomes were included in the analysis. Univariate Cox regression models were used in the training cohort to analyze risk factors associated with in patients with CSS. According to univariate Cox regression analysis, a multivariate Cox regression model was established to identify the independent risk factors for CSS.

Construction of Nomogram

The establishment of the nomogram mainly relies on the multivariate Cox regression model. We developed a new nomogram to predict CSS at 1-, 3-, and 5 years in elderly patients with ccRCC based on independent risk factors for univariate and multivariate screening. Because these independent risk factors are predictors of patient outcomes.

Validation of Nomogram

Subsequently, we used the consistency index (C-index) to test the accuracy and discrimination of the prediction model in the training cohort and validation cohort. Calibration curves of 1,000 bootstrap samples were used to test the accuracy of the prediction model. The area under the receiver operating curve (AUC) was used to test the model's prediction accuracy at 1-, 3-, and 5 years.

Clinical Utility

We used the decision analysis curve (DCA) to validate the clinical potential of the predictive model. DCA is a new algorithm for calculating net benefit under different thresholds and is currently most commonly used to evaluate the clinical application value of models. In addition, we divided patients into high-risk and low-risk groups based on each patient's nomogram score and cutoff values determined by the receiver operating curve. Log-rank test and Kaplan-Meier (K-M) curve were used to compare the survival difference between the two groups. In addition, we analyzed surgical differences between high-risk and low-risk patients based on risk grouping.

Statistical Analysis

Continuous measurement data were described by means and variance, and comparison between groups was performed by Chi-square test or non-parametric *U* test. Count data were characterized by frequency (%), and a chi-square test was used to compare groups. Univariate and multivariate Cox proportional regression models analyzed the survival and prognostic factors. Log-rank test and Kaplan-Meier (K-M) curve were used to analyze the survival differences between the patients. SPSS 26.0 and R Software 4.1.0 were used for all statistical analyses. Values of $p < 0.05$ were considered statistically significant.

RESULTS

Clinical Features

A total of 33,509 elderly patients with ccRCC were enrolled. All patients were randomly divided into the training cohort (N = 23,412) and the validation cohort (N = 10,097). The mean age of all patients was 72.7 ± 5.91 years, 29,034 (86.6%) white, 20,201 (60.3%) men, and 21,239 (63.4%) married. Among all patients, there were 16,447 (49.1%) tumors on the left side, 4,077 (12.2%) tumors at grade I, 17,480 (52.2%) tumors at grade II, 9,595 (28.6%) tumors at grade III, and 2,357 (7.03%) tumors at grade IV. The mean tumor size was 51.8 ± 32.3 mm, the T stage T1a had 14,019 (41.8%), T1b had 8,063 (24.1%), T2 had 2,987 (8.91%), T3 had 8,370 (25.0%), and T4 had 70 (0.21%). There were 32,620 (97.3%) patients in stage N0 and 31,320 (93.5%) patients in stage M0. There were 1,227 (3.66%) patients without surgery, 1,718 (5.13%) patients with local tumor excision, 8,791 (26.2%) patients with PN, and 21,773 (65.0%) patients with RN. There were 1,505 (4.49%) patients who received chemotherapy and 640 (1.91%) patients who received radiotherapy. The clinicopathological

information of the patients is shown in **Table 1**, and there was no significant difference between the training cohort and the validation cohort.

Univariate and Multivariate Cox Regression Analyses

Univariate Cox regression model was used to analyze all variables to screen for variables associated with CSS in elderly patients with ccRCC. Age, sex, race, marriage, year of diagnosis, tumor size, histological grade, TNM stage, and surgery were prognostic factors related to patient survival. We included these factors in a multivariate Cox regression analysis. We found that age, sex, race, marriage, tumor size, histological grade, TNM stage, and surgery were independent risk factors for CSS in elderly patients with ccRCC. In other words, these independent risk factors can be used as factors in predictive models to predict CSS of the patients. The univariate and multivariate analyses results are shown in **Table 2**.

Nomogram Construction for 1-, 3-, and 5-Year CSS

We constructed a nomogram to predict CSS in elderly patients with ccRCC (**Figure 2**). Tumor size was the most significant risk factor for in patients with CSS, as shown in the nomogram. The larger the tumor, the higher the patient's risk of death. Second, surgery is the second most important factor affecting patient survival. Survival was most increased in patients with PN, followed by RN and local tumor excision and was lowest in patients who did not undergo surgery. Tumor TNM stage and histological tumor grade are important risk factors for patient prognosis. Patients with higher TNM stage or histological grade had lower survival rates. In addition, the age of the patients was also a key factor; the older the patient, the greater the risk of death. Finally, married patients, women, and other races had a relatively high survival rate.

Validation of the Nomogram

In the training cohort and validation cohort, we first use the C-index to validate the accuracy of the nomogram. The C-index of the training cohort and validation cohort was 0.81 (95% CI: 0.802–0.818) and 0.818 (95% CI: 0.806–0.830), respectively, indicating the accuracy and discrimination of the nomogram. In the training cohort and validation cohort, the calibration curve showed that the observed value of the prediction model was highly consistent with the predicted value, proving that the prediction model had good accuracy (**Figure 3**). The AUC of the training cohort and validation cohort also suggested that the prediction model had good accuracy (**Figure 4**). In the training cohort, the AUC of the prediction model was 86.8 (95% CI, 85.5–88.0), 84.8 (95% CI, 83.8–85.9), and 82.8 (95% CI, 80.8–84.8) in 1-, 3-, and 5 years, respectively. In the validation cohort, the AUC of the prediction model was 86.0 (95% CI, 84.0–88.0), 85.0 (95% CI, 83.5–86.4), and 83.3 (95% CI, 81.6–85.0), respectively.

TABLE 1 | Clinicopathological characteristics of elderly patients with ccRCC.

	All	Training cohort	validation cohort	
	N = 33,509	*N* = 23,412	*N* = 10,097	*p*
Age	72.7 (5.91)	72.7 (5.90)	72.8 (5.95)	0.515
Race				0.073
white	29,034 (86.6%)	20,233 (86.4%)	8,801 (87.2%)	
black	2,030 (6.06%)	1,463 (6.25%)	567 (5.62%)	
other	2,445 (7.30%)	1,716 (7.33%)	729 (7.22%)	
Sex				0.313
Male	20,201 (60.3%)	14,156 (60.5%)	6,045 (59.9%)	
Female	13,308 (39.7%)	9,256 (39.5%)	4,052 (40.1%)	
Marriage				0.332
No	12,270 (36.6%)	8,533 (36.4%)	3,737 (37.0%)	
Married	21,239 (63.4%)	14,879 (63.6%)	6,360 (63.0%)	
Year of diagnosis				0.651
2004–2010	13,295 (39.7%)	9,308 (39.8%)	3,987 (39.5%)	
2010–2018	20,214 (60.3%)	14,104 (60.2%)	6,110 (60.5%)	
Laterality				0.138
Left	16,447 (49.1%)	11,554 (49.4%)	4,893 (48.5%)	
Right	17,062 (50.9%)	11,858 (50.6%)	5,204 (51.5%)	
Grade				0.760
I	4,077 (12.2%)	2,834 (12.1%)	1,243 (12.3%)	
II	17,480 (52.2%)	12,212 (52.2%)	5,268 (52.2%)	
III	9,595 (28.6%)	6,698 (28.6%)	2,897 (28.7%)	
IV	2,357 (7.03%)	1,668 (7.12%)	689 (6.82%)	
T				0.266
T1a	14,019 (41.8%)	9,743 (41.6%)	4,276 (42.3%)	
T1b	8,063 (24.1%)	5,694 (24.3%)	2,369 (23.5%)	
T2	2,987 (8.91%)	2,113 (9.03%)	874 (8.66%)	
T3	8,370 (25.0%)	5,811 (24.8%)	2,559 (25.3%)	
T4	70 (0.21%)	51 (0.22%)	19 (0.19%)	
N				0.080
N0	32,620 (97.3%)	22,815 (97.5%)	9,805 (97.1%)	
N1	889 (2.65%)	597 (2.55%)	292 (2.89%)	
M				0.362
M0	31,320 (93.5%)	21,902 (93.6%)	9,418 (93.3%)	
M1	2,189 (6.53%)	1,510 (6.45%)	679 (6.72%)	
Tumor size	51.8 (32.3)	51.7 (31.0)	52.0 (35.0)	0.449
Surgery				0.426
No	1,227 (3.66%)	867 (3.70%)	360 (3.57%)	
Local tumor excision	1,718 (5.13%)	1,173 (5.01%)	545 (5.40%)	
Partial nephrectomy	8,791 (26.2%)	6,127 (26.2%)	2,664 (26.4%)	
Radical nephrectomy	21,773 (65.0%)	15,245 (65.1%)	6,528 (64.7%)	
Chemotherapy				0.038
No/Unknown	32,004 (95.5%)	22,397 (95.7%)	9,607 (95.1%)	
Yes	1,505 (4.49%)	1,015 (4.34%)	490 (4.85%)	
Radiation:				0.354
No/Unknown	32,869 (98.1%)	22,976 (98.1%)	9,893 (98.0%)	
Yes	640 (1.91%)	436 (1.86%)	204 (2.02%)	

TABLE 2 | Univariate and multivariate analyses of CSS in training cohort.

	Univariate			Multivariate		
	HR	**95%CI**	**P**	**HR**	**95%CI**	**P**
Age	1.04	1.03–1.04	<0.001	1.057	1.053–1.06	<0.001
Race						
white	Reference			Reference		
black	0.83	0.72–0.96	0.013	1.129	1.035–1.232	0.006
other	0.91	0.8–1.03	0.137	0.879	0.806–0.959	0.004
Sex						
Male	Reference			Reference		
Female	0.81	0.76–0.87	<0.001	0.834	0.796–0.874	<0.001
Marriage						
No	Reference			Reference		
Married	0.92	0.86–0.98	0.01	0.852	0.814–0.893	<0.001
Year of diagnosis						
2004–2010	reference					
2010–2018	0.91	0.85–0.97	0.006			
Laterality						
Left	Reference					
Right	1.03	0.97–1.1	0.32			
Grade						
I	Reference			Reference		
II	1.32	1.15–1.5	<0.001	1.037	0.965–1.113	0.32
III	2.7	2.36–3.08	<0.001	1.207	1.117–1.305	<0.001
IV	7.07	6.12–8.17	<0.001	1.797	1.626–1.987	<0.001
T						
T1a	Reference			Reference		
T1b	2.14	1.93–2.38	<0.001	1.176	1.105–1.251	<0.001
T2	4.56	4.06–5.11	<0.001	1.189	1.088–1.299	<0.001
T3	6.19	5.65–6.78	<0.001	1.388	1.294–1.488	<0.001
T4	26.49	16.75–41.89	<0.001	1.866	1.178–2.955	0.008
N						
N0	Reference			Reference		
N1	10.3	9.26–11.46	<0.001	2.023	1.82–2.249	<0.001
M						
M0	Reference			Reference		
M1	11.01	10.2–11.87	<0.001	3.049	2.83–3.285	<0.001
Tumor size	1.01	1.01–1.01	<0.001	1.003	1.002–1.004	<0.001
Surgery						
No	Reference			referEnce		
Local tumor excision	0.1	0.08–0.13	<0.001	0.421	0.368–0.481	<0.001
Partial nephrectomy	0.08	0.07–0.09	<0.001	0.282	0.253–0.314	<0.001
Radical nephrectomy	0.25	0.22–0.27	<0.001	0.361	0.329–0.396	<0.001

Clinical Application of the Nomogram

Decision curve analysis showed a good clinical application value (**Figure 5**). Moreover, DCA results showed that the DCA of the nomogram was significantly superior to the traditional TNM staging system in predicting patient survival. We divided the patients into the high-risk group (overall score ≤67.9) and the low-risk group (overall score > 67.9). The K-M curve showed that in the training cohort and validation cohort, the survival rate of patients in the high-risk group was significantly lower than that in the low-risk group (**Figure 6**). The 1-, 3-, and 5 years survival rates in high-risk group were 0.920 (95% CI, 0.916–0.924), 0.808 (95% CI, 0.802–0.815), and 0.732 (95% CI, 0.724–0.740), respectively. The 1-, 3-, and 5 years survival rates in the low-risk group were 0.992 (95% CI, 0.991–0.994), 0.976 (95% CI, 0.973–0.979), and 0.954 (95% CI, 0.950–0.957), respectively. Based on their risk grouping, we analyzed surgical options for patients in the low-risk and high-risk groups. In the low-risk group, we found that all patients received surgical treatment, and most

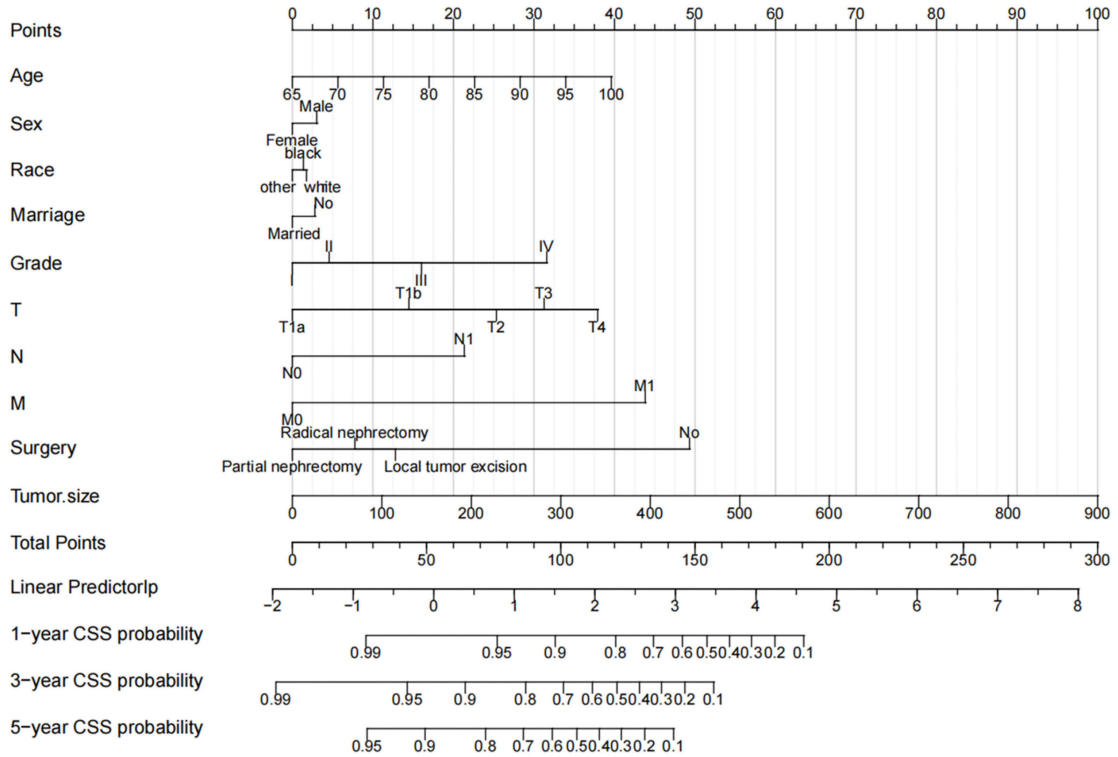


FIGURE 2 | The nomogram predicts CSS in elderly patients with ccRCC at 1-, 3-, and 5 years. CSS, cancer-specific survival; ccRCC, clear cell renal cell carcinoma.

patients received RN and PN, with significantly higher survival rates than local tumor excision. In the high-risk group, most patients who underwent RN had significantly higher survival rates than those who did not, with the highest survival rates among patients who underwent PN and local tumor excision (**Figure 7**).

Online Application of Nomogram

Based on our established programs, we developed a web application to predict 1-, 3-, 5 years CSS in elderly patients with ccRCC. Log on to https://liujiayan.shinyapps.io/DynNomapp/ can enter the website. The survival probability of patients can be obtained by inputting clinicopathological information of patients.

DISCUSSION

Renal cell carcinoma accounts for about 2% of all cancer diagnoses and deaths worldwide, with a higher incidence in developed countries and a severe burden on economic development. RCC is one of the top ten malignant tumors threatening human health (17). CcRCC is a category of RCC with a poor prognosis (7), and its risk factors are controversial. For example, Faramawi et al. (18) found that red meat consumption was an independent risk factor for the prognosis of ccRCC through meta-analysis, while Eun et al. (19) proved no clear correlation through 13 prospective studies. Similar conclusions also appeared when alcohol consumption (20) and exercise (21) were considered risk factors. Therefore, it is crucial to analyze and screen out specific risk factors, establish a predictive model, and accurately predict the prognosis of ccRCC patients.

Age is a critical factor in the occurrence and development of cancer, and so is ccRCC. With age, the probability of genetic mutations leading to cancer also increases. Aging is associated with highly reproducible DNA methylation changes, which helps to explain the higher prevalence of malignant tumors in the elderly (22). There is no consensus on defining the age of elderly patients, but more than 60% of initial cancer diagnoses and more than 70% of cancer deaths occur in patients over 65 years old (8). To improve the accuracy and representativeness of the prediction model, we included patients with ccRCC over 65 years old in this study. It is worth noting that newly diagnosed cases of ccRCC in the elderly are still mainly manifested by renal masses (8), and these masses usually show low-grade malignancy. However, clinicians mostly choose RN to avoid the poor prognosis caused by tumor metastasis, while such excessive treatment leads to more complications and even death in elderly patients (23, 24) and again highlights the crucial clinical value of this study.

Several studies have shown that married patients have increased cancer relative survival, even for untreated tumors (25–27). It may be related to the emotion of marriage and decision-making of medical treatment. Perini et al. (28) found

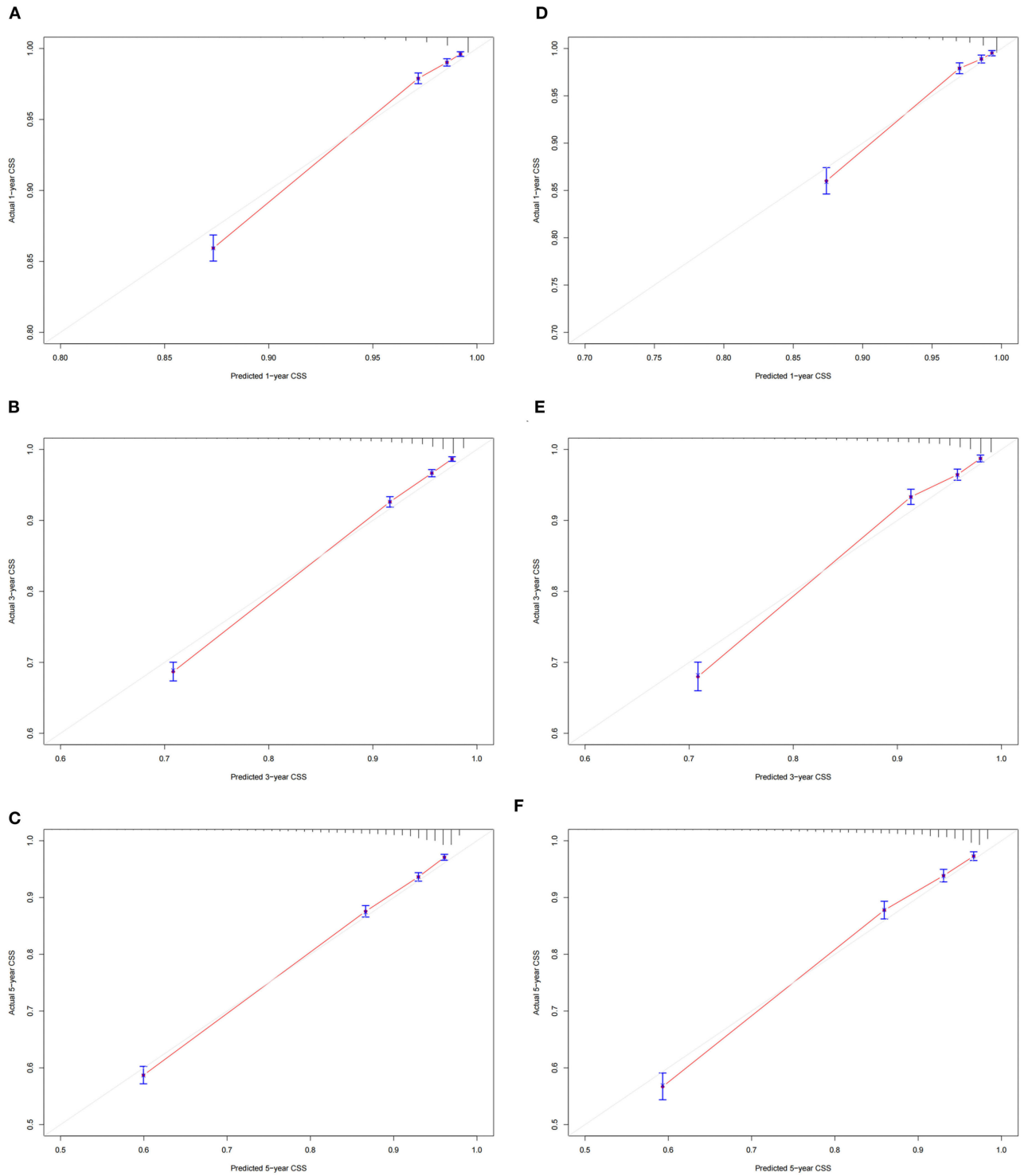


FIGURE 3 | Calibration curve of the nomogram. **(A–C)** Calibration curves of 1 -, 3 -, and 5 years CSS in the training cohort; **(D–F)** calibration curves of 1-, 3-, and 5 years CSS in the validation cohort. CSS, cancer-specific survival.

that depression was a risk factor for the low survival rate of cancer patients, while John et al. (29) found that the incidence of depression in men after marriage was significantly lower than that before marriage, which provided strong evidence for emotional factors. In addition, married cancer patients can detect physical abnormalities early and actively cooperate with

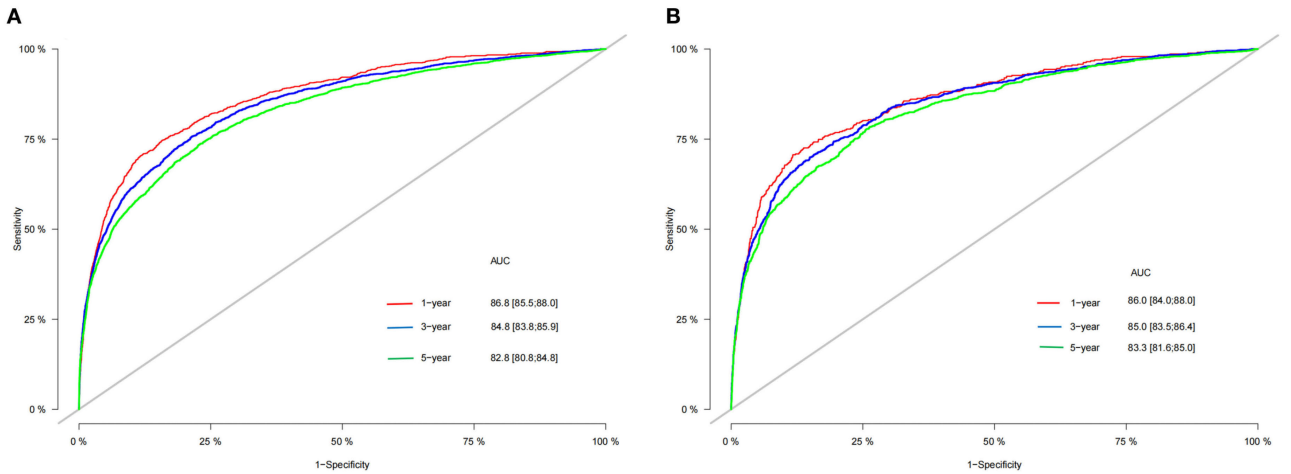


FIGURE 4 | AUC for predicting 1-, 3-, and 5 years CSS in training cohort **(A)** and validation cohort **(B)**. CSS, cancer-specific survival; AUC, area under the receiver operator curve.

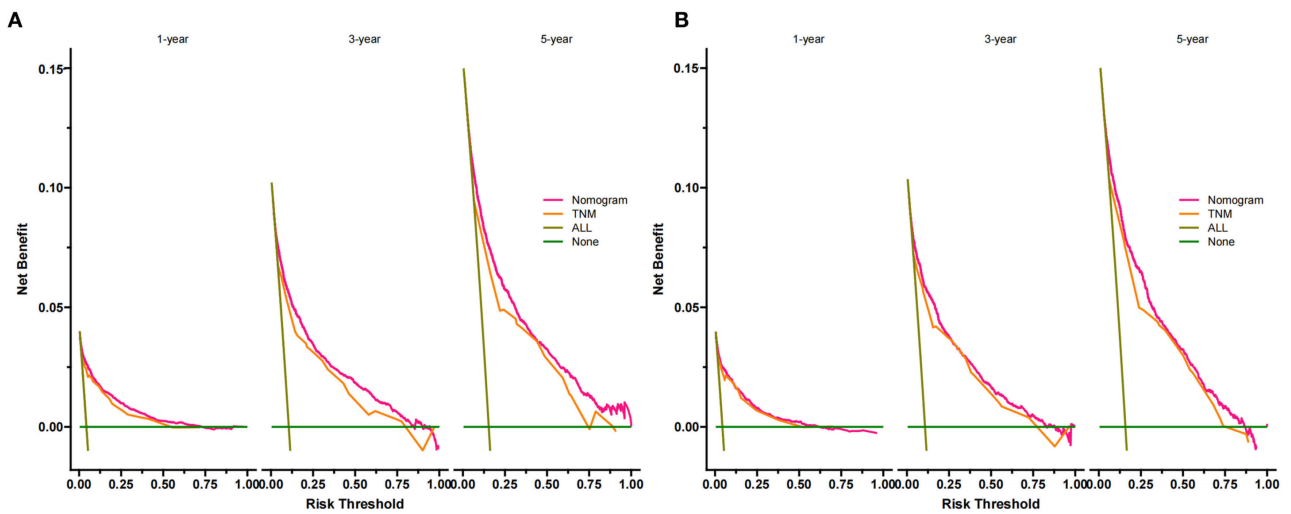


FIGURE 5 | DCA of the nomogram in the training cohort **(A)** and the validation cohort **(B)**. The Y-axis represents a net benefit, and the X-axis represents threshold probability. The green line means no patients died, and the dark green line means all patients died. When the threshold probability is between 0 and 100%, the net benefit of the model exceeds all deaths or none. DCA, decision curve analysis.

treatment, possibly out of the responsibility to the family after marriage (30). In conclusion, married patients showed significant improvement in RR compared with unmarried patients for cancer with a relatively low degree of malignancy, such as patients with ccRCC included in this study. Marriage is not beneficial for highly malignant cancers, such as pancreatic cancer and renal malignant rhabdomyoma tumor (31, 32).

On average, two individuals of the same sex are more than 99.9% alike. However, the genetic similarity between male and female individuals is only 98.5%(33). According to data from the SEER database, men tend to be diagnosed with RCC with higher stage and grade than women (34). Based on data analysis from the Multicenter Society for Renal Oncology Collaborative Study (CORONA) database, women's CSS (hazard ratio [HR] 0.75, $p < 0.001$) and overall survival (OS) (HR 0.80, $p < 0.001$) were significantly better than those of men (35). In this study, the prognosis of female patients with ccRCC was also substantially better than that of male patients (HR: 0.834, 95% CI: 0.796–0.874, $p < 0.001$).

Classical RN involves removing the kidney, perirenal adipose tissue, adrenal gland, and local lymph nodes. It became the preferred treatment for ccRCC for some time. However, the massive loss of nephron caused by surgery can induce chronic kidney disease and cause death (36). Nevertheless, PN can preserve normal nephron to the maximum extent and is widely used in ccRCC cases with tumor sizes <7 cm in length. Nowadays, the biggest challenge for PN is to avoid positive excision margin and reduce postoperative complications, such

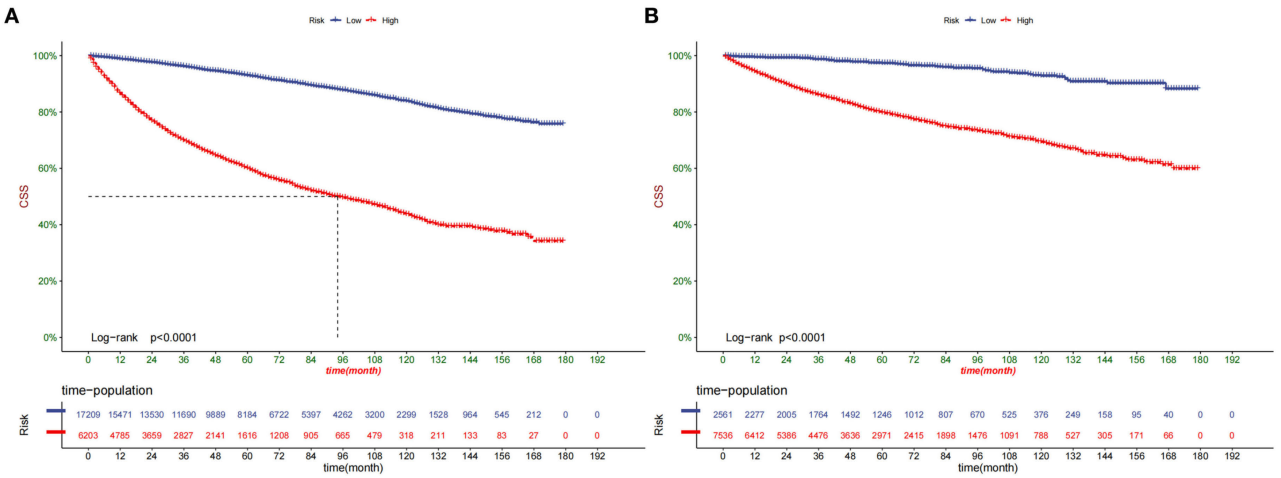


FIGURE 6 | Kaplan-Meier curves of patients in the low-risk and high-risk groups in training cohort **(A)** and validation cohort **(B)**.

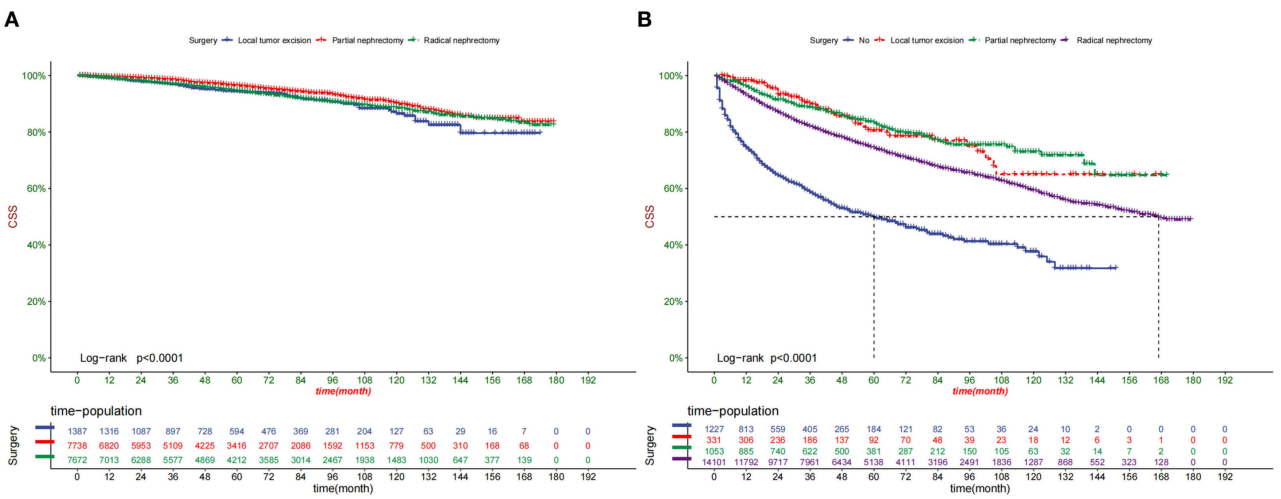


FIGURE 7 | Kaplan-Meier curves of patients with different surgery in the low-risk group **(A)** and high-risk group **(B)**.

as hematuria, perirenal hematoma, and urinary fistula (37). Local tumor excision and active monitoring (AS) represented by RFA have been widely used in elderly patients and achieved good results due to contraindicated comorbidity and short-life expectancy. In a meta-analysis of case studies, Dib et al. (38) showed that cryoablation and RFA efficacy were 89 and 90%, respectively. In addition, RFA and other treatments can be operated under the guidance of ultrasound or CT positioning, effectively avoiding the damage of anesthetics to the heart, liver, and other organs of the elderly. Nevertheless, the surgical recurrence rate of local tumor resection is much higher than that of RN and PN (37). This study found that low-risk patients were more likely to benefit from PN and RN, while high-risk patients were more likely to benefit from PN and local tumor excision.

TNM staging of RCC was based on the 2021 European Association of Urology (EAU) updated criteria (39). The Fuhrman grading system was adopted for tumor classification. According to WHO recommendations, grade I was defined as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated (36). We found that high TNM staging and histologic grade were associated with poor prognosis, which is consistent with the conclusions of previous studies (40).

In recent years, the broad application of the nomogram prediction model has gradually replaced the traditional TNM staging. First, the c-index of traditional TMN staging for prognosis prediction is about 0.6, which is relatively inaccurate. Second, the prognosis of patients is also affected by many non-anatomic factors, such as age, gender, race, marital status, and surgical methods (41). However, the C-index of this study is 0.81, which is of higher accuracy than the traditional TNM system. Nomogram is a data-based graphical computing tool that can estimate the risk of developing a disease based on staging systems, such as the American Joint Commission on Cancer (AJCC) and

other key risk factors related to prognosis (42). Today, many kinds of nomograms have been developed and applied to RCC, such as Chen et al. (43), who included 6,105 elderly patients with metastatic RCC in the SEER database from 2010 to 2015, indicating the advantages of PN in elderly mRCC. Michele et al. (44) included 4,541 patients with RCC over 75 years of age in the SEER database from 2004 to 2014 and proved that PN did not increase the mortality of elderly patients. With the update of the SEER database and the improvement of patients' clinical information, this study included 33,509 elderly CCRCC patients from 2004 to 2018, which was superior to previous studies on elderly RCC in terms of the number of cases and c-index accuracy.

Still, there are some limitations to our study. First of all, our study was retrospective, and selection bias was inevitable. Secondly, some variables, such as patients' body mass index (BMI), smoking, drinking, heart disease, and other data cannot be obtained from the SEER database. However, we included significant variables, such as tumor stage, surgery, age, and other vital factors. Finally, our prediction model is only validated internally, and further external validation is necessary to validate the model's accuracy.

CONCLUSION

In this study, we explored prognostic factors in elderly patients with ccRCC. We found that age, sex, marriage, TNM stage, surgery, and tumor size were independent risk factors for CSS. We constructed a new nomogram to predict CSS in elderly patients with ccRCC with good accuracy and reliability, providing clinical guidance for patients and physicians.

DATA AVAILABILITY STATEMENT

The data analyzed in this study was obtained from the Surveillance, Epidemiology, and End Results (SEER) Program, the following licenses/restrictions apply: To request access to the SEER Incidence Databases, you will first need to create a SEER*Stat Account and submit a request form. Requests to access these datasets should be directed to: https://service.cancer.gov/seer-data-access. For questions regarding the data access request process, send an email to SEERDataAccess@mail.nih.gov.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/ participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JW and CZ designed the study. CZ, JW, JL, LJ, ML, and XT collected and analyzed the data. JW drafted the initial manuscript. CZ, TM, JL, ZZ, and DH revised the article critically. CZ, JL, DH, LJ, and XT reviewed and edited the article. All authors approved the final manuscript.

FUNDING

This study was supported by the Kunming City Health Science and Technology Talent 1000 training Project [no. 2020-SW (Reserve)-112], Kunming Health and Health Commission Health Research Project (no. 2020-0201-001), Kunming Xishan District Science and Technology Project (no. 2020-Xike word 23), Kunming Medical Joint Project of Yunnan Science and Technology Department (no. 202001 AY070001-271), and Special Key Project of Chongqing Technology Innovation and Application Development (no. Cstc2019jscx-tjsbX0003).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [published correction appears in CA. *Cancer J Clin*. (2020). 70:313. doi: 10.3322/caac.21609
2. Quivy A, Daste A, Harbaoui A, Duc S, Bernhard JC, Gross-Goupil M, et al. Optimal management of renal cell carcinoma in the elderly: a review. *Clin Interv Aging*. (2013) 8:433–42. doi: 10.2147/CIA.S30765
3. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, et al. Renal cell carcinoma. *Nat Rev Dis Primers*. (2017). 3:17009. doi: 10.1038/nrdp.2017.9
4. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. [published correction appears in CA. *Cancer J Clin*. (2021). 71:359. doi: 10.3322/caac.21654
5. Kim H, Shim BY, Lee SJ, Lee JY, Lee HJ, Kim IH. Loss of Von Hippel-Lindau (VHL) Tumor Suppressor Gene Function: VHL-HIF Pathway and Advances in Treatments for Metastatic Renal Cell Carcinoma (RCC). *Int J Mol Sci*. (2021) 22:9795. doi: 10.3390/ijms22189795
6. Jonasch E, Gao J, Rathmell WK. Renal cell carcinoma. *BMJ*. (2014) 349:g4797. doi: 10.1136/bmj.g4797
7. Jonasch E, Walker CL, Rathmell WK. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. *Nat Rev Nephrol*. (2021) 17:245–61. doi: 10.1038/s41581-020-00359-2
8. González León T, Morera Pérez M. Renal cancer in the elderly. *Curr Urol Rep*. (2016) 17:6. doi: 10.1007/s11934-015-0562-2
9. Motzer RJ, Banchereau R, Hamidi H, Powles T, McDermott D, Atkins MB, et al. Molecular Subsets in Renal Cancer Determine Outcome to Checkpoint and Angiogenesis Blockade. *Cancer Cell*. (2020). 38:803–17.e4. doi: 10.1016/j.ccell.2020.10.011
10. Labriola MK, Zhu J, Gupta RT, McCall S, Jackson J, Kong EF, et al. Characterization of tumor mutation burden, PD-L1 and DNA repair genes to assess relationship to immune checkpoint inhibitors response in metastatic renal cell carcinoma. *J Immunother Cancer*. (2020) 8:e000319. doi: 10.1136/jitc-2019-000319
11. Murakami T, Tanaka N, Takamatsu K, Hakozaki K, Fukumoto K, Masuda T, et al. Multiplexed single-cell pathology reveals the association of CD8 T-cell heterogeneity with prognostic outcomes in renal cell carcinoma. *Cancer Immunol Immunother*. (2021) 70:3001–13. doi: 10.1007/s00262-021-03006-2
12. Haddad H, Rini BI. Current treatment considerations in metastatic renal cell carcinoma. *Curr Treat Options Oncol*. (2012) 13:212–29. doi: 10.1007/s11864-012-0182-8

13. Parker WP, Cheville JC, Frank I, Zaid HB, Lohse CM, Boorjian SA, et al. Application of the stage, size, grade, and necrosis (SSIGN) score for clear cell renal cell carcinoma in contemporary patients. *Eur Urol.* (2017) 71:665–73. doi: 10.1016/j.eururo.2016.05.034
14. Beisland C, Gudbrandsdottir G, Reisæter LA, Bostad L, Wentzel-Larsen T, Hjelle KM. Contemporary external validation of the Leibovich model for prediction of progression after radical surgery for clear cell renal cell carcinoma. *Scand J Urol.* (2015) 49:205–10. doi: 10.3109/21681805.2014.980844
15. Awais M, Ghayvat H, Krishnan Pandarathodiyil A, Nabillah Ghani WM, Ramanathan A, Pandya S, et al. Healthcare professional in the loop (HPIL): classification of standard and oral cancer-causing anomalous regions of oral cavity using textural analysis technique in autofluorescence imaging. *Sensors (Basel).* (2020) 20:5780. doi: 10.3390/s20205780
16. Shah A, Ahirrao S, Pandya S, Kotecha K, Rathod S. Smart cardiac framework for an early detection of cardiac arrest condition and risk. *Front Public Health.* (2021) 9:762303. doi: 10.3389/fpubh.2021.762303
17. Sharma R. An examination of colorectal cancer burden by socioeconomic status: evidence from GLOBOCAN (2018) *EPMA J.* (2019). 11:95–117. doi: 10.1007/s13167-019-00185-y
18. Faramawi MF, Johnson E, Fry MW, Sall M, Zhou Y. Consumption of different types of meat and the risk of renal cancer: meta-analysis of case-control studies [published correction appears in Cancer Causes Control. *Cancer Causes Control*. (2007). 18:125–33. doi: 10.1007/s10552-006-0104-9
19. Lee JE, Spiegelman D, Hunter DJ, Albanes D, Bernstein L, van den Brandt PA, et al. Fat, protein, and meat consumption and renal cell cancer risk: a pooled analysis of 13 prospective studies. *J Natl Cancer Inst.* (2008) 100:1695–706. doi: 10.1093/jnci/djn386
20. Cohortiawan VW, Stram DO, Nomura AM, Kolonel LN, Henderson BE. Risk factors for renal cell cancer: the multiethnic cohort. *Am J Epidemiol.* (2007) 166:932–40. doi: 10.1093/aje/kwm170
21. Bergström A, Terry P, Lindblad P, Lichtenstein P, Ahlbom A, Feychting M, et al. Physical activity and risk of renal cell cancer. *Int J Cancer.* (2001) 92:155–7. doi: 10.1002/1097-0215(200002)9999:9999<::AID-IJC1162>3.0.CO;2-S
22. Lin Q, Wagner W. Epigenetic aging signatures are coherently modified in cancer. *PLoS Genet*. (2015) 11:e1005334. doi: 10.1371/journal.pgen.1005334
23. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol.* (2019) 30:706–720. doi: 10.1093/annonc/mdz056
24. Gontero P, Joniau S, Oderda M, Ruutu M, Van Poppel H, Laguna MP, et al. Active surveillance for small renal tumors: have clinical concerns been addressed so far? *Int J Urol.* (2013) 20:356–61. doi: 10.1111/j.1442-2042.2012.03227.x
25. Li Z, Wang K, Zhang X, Wen J. Marital status and survival in patients with rectal cancer: A population-based STROBE cohort study [published correction appears in Medicine (Baltimore). (2018). 97:e10872. *Medicine (Baltimore).* (2018) 97:e0637. doi: 10.1097/MD.0000000000010637
26. Gomez SL, Hurley S, Canchola AJ, Keegan TH, Cheng I, Murphy JD, et al. Effects of marital status and economic resources on survival after cancer: A population-based study. *Cancer.* (2016) 122:1618–25. doi: 10.1002/cncr.29885
27. Martínez ME, Anderson K, Murphy JD, Hurley S, Canchola AJ, Keegan TH, et al. Differences in marital status and mortality by race/ethnicity and nativity among California cancer patients. *Cancer.* (2016) 122:1570–8. doi: 10.1002/cncr.29886
28. Bortolato B, Hyphantis TN, Valpione S, Perini G, Maes M, Morris G, et al. Depression in cancer: The many biobehavioral pathways driving tumor progression. *Cancer Treat Rev.* (2017) 52:58–70. doi: 10.1016/j.ctrv.2016.11.004
29. St John PD, Montgomery PR. Marital status, partner satisfaction, and depressive symptoms in older men and women. *Can J Psychiatry.* (2009) 54:487–92. doi: 10.1177/070674370905400710
30. Enewold L, Harlan LC, Tucker T, McKenzie S. Pancreatic cancer in the USA: persistence of undertreatment and poor outcome. *J Gastrointest Cancer.* (2015) 46:9–20. doi: 10.1007/s12029-014-9668-x
31. Zhanghuang C, Chen S, Li L, Yang Z, Xie Y, Li J, et al. Clinical and molecular differentiation between malignant rhabdoid tumor of the kidney and normal tissue: a two-case report. *Front Oncol.* (2021) 11:659709. doi: 10.3389/fonc.2021.659709
32. Merrill RM, Johnson E. Benefits of marriage on relative and conditional relative cancer survival differ between males and females in the USA. *J Cancer Surviv.* (2017) 11:578–89. doi: 10.1007/s11764-017-0627-y
33. Lucca I, Klatte T, Fajkovic H, de Martino M, Shariat SF. Gender differences in incidence and outcomes of urothelial and kidney cancer. *Nat Rev Urol.* (2015) 12:653. doi: 10.1038/nrurol.2015.232
34. Aron M, Nguyen MM, Stein RJ, Gill IS. Impact of gender in renal cell carcinoma: an analysis of the SEER database. *Eur Urol.* (2008) 54:133–40. doi: 10.1016/j.eururo.2007.12.001
35. May M, Aziz A, Zigeuner R, Chromecki T, Cindolo L, Schips L, et al. Gender differences in clinicopathological features and survival in surgically treated patients with renal cell carcinoma: an analysis of the multicenter CORONA database. *World J Urol.* (2013) 31:1073–80. doi: 10.1007/s00345-013-1071-x
36. Borgmann H, Musquera M, Haferkamp A, Vilaseca A, Klatte T, Shariat SF, et al. Prognostic significance of Fuhrman grade and age for cancer-specific and overall survival in patients with papillary renal cell carcinoma: results of an international multi-institutional study on 2189 patients. *World J Urol.* (2017) 35:1891–7. doi: 10.1007/s00345-017-2078-5
37. MacLennan S, Imamura M, Lapitan MC, Omar MI, Lam TB, Hilvano-Cabungcal AM, et al. Systematic review of perioperative and quality-of-life outcomes following surgical management of localised renal cancer. *Eur Urol.* (2012) 62:1097–117. doi: 10.1016/j.eururo.2012.07.028
38. El Dib R, Touma NJ, Kapoor A. Cryoablation vs radiofrequency ablation for the treatment of renal cell carcinoma: a meta-analysis of case series studies. *BJU Int.* (2012) 110:510–6. doi: 10.1111/j.1464-410X.2011.10885.x
39. Rouprêt M, Babjuk M, Burger M, Capoun O, Cohen D, Compérat EM, et al. European Association of Urology Guidelines on Upper Urinary Tract Urothelial Carcinoma: (2020). *Update Eur Urol.* (2021) 79:62–79. doi: 10.1016/j.eururo.2020.05.042
40. Hofbauer SL, de Martino M, Seemann C, Zamani N, Lucca I, Haitel A, et al. Associations between presenting symptoms, clinicopathological parameters, and prognosis in a contemporary series of patients with renal cell carcinoma. *Korean J Urol.* (2014) 55:505–10. doi: 10.4111/kju.2014.55.8.505
41. Serdy KM, Leone JP, Dabbs DJ, Bhargava R. Male breast cancer. *Am J Clin Pathol.* (2017) 147:110–9. doi: 10.1093/ajcp/aqw207
42. Liu X, Wu Z, Lin E, Li W, Chen Y, Sun X, et al. Systemic prognostic score and nomogram based on inflammatory, nutritional and tumor markers predict cancer-specific survival in stage II-III gastric cancer patients with adjuvant chemotherapy. *Clin Nutr.* (2019) 38:1853–60. doi: 10.1016/j.clnu.2018.07.015
43. Chen J, He Q, Liu W, Li Y, Zhuang W. The effect of cytoreductive partial nephrectomy in elderly patients with metastatic renal cell carcinoma. *Clin Interv Aging.* (2020) 15:431–9. doi: 10.2147/CIA.S243902
44. Marchioni M, Preisser F, Bandini M, Nazzani S, Tian Z, Kapoor A, et al. Comparison of partial versus radical nephrectomy effect on other-cause mortality, cancer-specific mortality, and 30-day mortality in patients older than 75 years. *Eur Urol Focus.* (2019) 5:467–73. doi: 10.1016/j.euf.2018.01.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhanghuang, Wang, Zhang, Jin, Tan, Mi, Liu, Li and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fpubh.2021.813641

Cognitive Impairment of Patient With Neurological Cerebrovascular Disease Using the Artificial Intelligence Technology Guided by MRI

Lifang Zhang[1,2*†], *Yanran Li*[3†], *Lin Bian*[1], *Qingrong Luo*[1], *Xiaoxi Zhang*[2] *and Bing Zhao*[1]

[1] Department of Neurology, Changzhi People's Hospital, Changzhi Medical College, Changzhi, China, [2] Department of Mental Health, Changzhi Medical College, Changzhi, China, [3] Department of Radiology, First Affiliated Hospital of Xinjiang Medical University, Ürümqi, China

This study was to explore the application of MRI based on artificial intelligence technology combined with neuropsychological assessment to the cognitive impairment of patients with neurological cerebrovascular diseases. A total of 176 patients were divided into a control group, a vascular cognitive impairment non-dementia (VCIND) group, a vascular dementia (VD) group, and an Alzheimer's disease (AD) group. All patients underwent MRI and neuropsychological evaluation and examination, and an improved fuzzy C-means (FCM) clustering algorithm was proposed for MRI processing. It was found that the segmentation accuracy (SA) and similarity (KI) data of the improved FCM algorithm used in this study were higher than those of the standard FCM algorithm, bias-corrected FCM (BCFCM) algorithm, and rough FCM (RFCM) algorithm ($p < 0.05$). In the activities of daily living (ADL), the values in the VCIND group (23.55 ± 6.12) and the VD group (28.56 ± 3.1) were higher than that in the control group (19.17 ± 3.67), so the hippocampal volume was negatively correlated with the ADL ($r = -0.872$, $p < 0.01$). In the VCIND group (52.4%), VD group (31%), and AD group (26.1%), the proportion of patients with the lacunar infarction distributed on both sides of the brain and the number of multiple cerebral infarction lesions (76.2, 71.4, and 71.7%, respectively) were significantly higher than those in the control group (23.9 and 50%). In short, the improved FCM algorithm showed a higher segmentation effect and SA for MRI of neurological cerebrovascular disease. In addition, the distribution, number, white matter lesions, and hippocampal volume of lacunar cerebral infarction were related to the cognitive impairment of patients with cerebrovascular diseases.

Keywords: artificial intelligence technology, cognitive impairment, MRI, neuropsychological assessment, neurological cerebrovascular disease

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Sharnil Pandya, Symbiosis International University, India
Rutvij Jhaveri, Pandit Deendayal Petroleum University, India
Mohammad Kamrul Hasan, National University of Malaysia, Malaysia

****Correspondence:***
Lifang Zhang
zlf_7296@163.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *12 November 2021*
Accepted: *06 December 2021*
Published: *03 March 2022*

Citation:
Zhang L, Li Y, Bian L, Luo Q, Zhang X and Zhao B (2022) Cognitive Impairment of Patient With Neurological Cerebrovascular Disease Using the Artificial Intelligence Technology Guided by MRI. Front. Public Health 9:813641. doi: 10.3389/fpubh.2021.813641

INTRODUCTION

Cerebrovascular disease is one of the main factors to cause vascular cognitive impairment (VCI) (1, 2). With the aging of current society, the incidence of cerebrovascular disease and VCI has increased significantly, ranking second only to Alzheimer's disease (AD), and is an important cause of the cognitive decline in the elderly (3). VCI is a common clinical type. It is the same as other cognitive diseases, and its clinical characteristics are gait, mood, behavior, and urination disorders (4). When the early cognitive impairment occurs, the condition will be mild so that it will often be ignored by others. When it develops into a disorder or even dementia, it is already too late and is not easy to diagnose and treat (5, 6). Therefore, it is very important to detect and find the vascular dysfunction as early as possible.

At present, the diagnosis of VCI is mainly based on the neuropsychological assessment to measure the cognitive function of the patients and to determine the relationship between cognitive impairment and brain disease. If the computer MRI is used as an auxiliary diagnostic method, the diagnosis results will be more reliable (7). With the continuous increase in the breadth and depth of clinical applications of medical images, medical images have become one of the important research directions in medical image processing. The technical difficulty of the brain MRI segmentation is mainly reflected in the following aspects. (1) Due to the principle and technical limitations of MRI and the influence of various factors in the image acquisition process, the image has partial volume effects, field offset effects, noise, motion artifacts, etc. (8, 9). (2) The tissue structure of the brain is complex. Normal tissues in the brain MRI mainly include the cerebral cortex, gray matter, white matter, cerebrospinal fluid, etc.; and the diseased tissues mainly include tumors, edema, and necrosis and cystic transformation within the tumors, and each type of tissue structure is newly transmitted and complex. (3) The space occupation, new transmission, and size of intracranial lesions are uncertain; the gray distribution between different soft tissues or between the diseased tissue and the soft tissue is uneven; and the density distribution is aliased, which is very convenient and very fuzzy. (4) There are big differences in the organizational results of different individuals, even the same person at different ages will have great differences (10–12). Therefore, the automatic segmentation technology using brain tissue MRI plays an important role in the diagnosis and treatment of brain diseases. MRI is the most widely used technique in radio imaging. As a dynamic and flexible technique, MRI can use different parameters for imaging, including longitudinal relaxation time (T1), lateral relaxation time (T2), T1 weighting imaging (T1WI), T2 weighting imaging (T2WI), etc. (13, 14). The imaging of these different parameters can effectively use brain MRI segmentation. Therefore, this study adopts an improved fuzzy C-means (FCM) clustering algorithm to overcome the shortcomings of FCM itself. The clustering-based segmentation method is used to divide the image pixels into several categories according to their similarity. The distance between different individuals and the distance between classes are larger, and the distance between individuals of the same class is smaller. FCM assigns each pixel to each type of organization with a different degree of membership and then has a certain criterion to uniquely divide each pixel into a certain type of organization to obtain the result of segmentation (15, 16).

Therefore, the relevant case data and MRI data of the patients in the Department of Neurology and Physical Examination, Changzhi People's Hospital are selected. In this study, lacunar cerebral infarction, white matter lesions, and hippocampal volume of patients with neurological cerebrovascular disease were studied. Through MRI based on artificial intelligence technology, the neuropsychological evaluation was combined to analyze its correlation with cognitive impairment, aiming to provide a reliable theoretical basis for early diagnosis and the detection, diagnosis, and treatment of vascular cognitive impairment.

MATERIALS AND METHODOLOGY

Information of Research Objects

The relevant case data and MRI data of the patients in the Department of Neurology and Physical Examination, Changzhi People's Hospital from October 1, 2018 to February 1, 2020 were selected. The studies involving human participants were reviewed and approved by Changzhi People's Hospital Ethics Committee. The patients/participants provided their written informed consent to participate in this study. The patients were divided into a control group, a vascular cognitive impairment non-dementia (VCIND) group, a vascular dementia (VD) group, and an AD group according to the Clinical Dementia Rating (CDR) Scale. The classification and diagnostic criteria of VD in the vascular cognitive impairment in China in 2019 were met in the VD group, and the core clinical criteria in the diagnosis of AD revised by the NIA-AA in 2011 were met in the AD group (13). There were 42, 42, and 46 cases in the VCIND, VD, and AD groups, respectively. About 46 cases of healthy patients without memory loss in the Department of Physical Examination were included in the control group, and they had no other related diseases that may affect their cognitive function.

The inclusion criteria: first, completed clinical case and MRI data; second, no acute cerebrovascular disease in at least 1 year; third, meeting the criteria of VD and AD in the diagnostic guidelines for the diagnosis and treatment of vascular cognitive impairment in China Diagnostic criteria for vascular VCI; and forth, no cerebral white matter lesions caused by hydrocephalus, cerebral hemorrhage shown in the MRI.

Exclusion criteria: first, more than 50% of the cranial vascular narrowing shown in the intracranial vascular ultrasonography, head MRI, or head CT angiography; second, having abnormal neurological diseases such as Parkinson's, brain tumors, and encephalitis; third, moderate aphasia, depression, or mental disorders, and other case being unable to complete the investigation; and forth, hereditary small blood vessel disease.

Neuropsychological Assessment

The patients were performed with relevant questions and answers for the test and evaluation in a quiet environment. The test tables included the mini-mental state examination (MMSE), montreal cognitive assessment (MOCA), memory and executive screening (MES), activities of daily living (ADL), and CDR (14). The

neuropsychological assessment was evaluated by an experienced neurologist in the neuropsychological assessment.

MRI Examination Based on the Artificial Intelligence Technology

The MRI examination: the Siemens verio3.0T magnetic resonance scanner was used to scan the patient's head for conventional transverse, axial T2WI, sagittal T1WI, coronary T2 fluid attenuation inversion recovery sequence (FLAIR), and hippocampal volume scan. The results of the MRI examination were identified by 2–3 imaging experienced physicians. For patients being unable to cooperate with the MRI examination temporarily, oral chloral hydrate could be used for sedation.

Improved FCM Image Segmentation Algorithm

The correct segmentation of brain tissue is an important prerequisite for the location and description of lesions. Noise and offset fields, as important factors affecting the quality of brain MRI, often exist at the same time. Therefore, image segmentation algorithms that can remove noise and offset fields are very important for the correct division of brain tissue (17). Therefore, based on the FCM algorithm, an MRI segmentation algorithm that removed noise and corrected the offset field was proposed.

The actual image would be affected by additive noise and multiplicative offset field, which can be expressed as follows:

$$q_i = a_i p_i + c \tag{1}$$

In Equation (1), q_{i} represents a noisy image, a_i represents an offset field, p_i represents an uncontaminated image, and c represents noise. In actual operation, both the offset field and the uncontaminated image are unknown parameters, so the offset field was regarded as a slowly changing and smooth uneven field.

The traditional coherent local intensity clustering (CLIC) algorithm based on the gray correction can be used to evaluate the local neighborhood M_{i} of the offset field, then below equation could be obtained:

$$M_i = \left\{ j : |j - i| \le r \right\} \tag{2}$$

In Equation (2), r represents the radius of the neighborhood. Taking into account the slow change of the offset field, a_j can be replaced by a_i in a small range, so the q_j belonging to the l-th category can be expressed as follows:

$$q_j = a_j d_l + c \tag{3}$$

The cost function of pixel q_i can be expressed as follows:

$$\lambda_i = \sum_{l=1}^{m} \sum_{j \epsilon M_i} \mu_{lj}^{m} K(i-j) \left\| q_i - a_i d_l \right\|^2 \tag{4}$$

In Equation (4), λ_i represents the cost function, and $K(i - j)$ represents the weight coefficient, and Equation (5) could be obtained:

$$K\left(i-j\right) = \begin{cases} \frac{e^{-|i-j|^2}}{v} / 2\sigma^2 & |i-j| \le r \\ 0 & other \end{cases} \tag{5}$$

The total cost function of the CLIC algorithm can be combined with the FCM algorithm to get the equation below.

$$T_c = \sum_{i=1}^{m} \sum_{l=1}^{d} \sum_{j \epsilon M_i} \mu_{lj}^{m} K(i-j) \left\| q_i - a_i d_l \right\|^2 \tag{6}$$

In Equation (6), T_C is the total cost function. The CLIC algorithm can directly process the original data, and the obtained offset field is smooth and changes slowly, without subsequent filtering operations. However, because the objective function does not consider the constraints of domain information, it is more sensitive to noise parameters. In this paper, the spatial information is added to the objective function of the CLIC algorithm, and a new weight coefficient R_{li} is proposed, which is as follows:

$$T_{CR} = \sum_{i=1}^{m} \sum_{l=1}^{d} \sum_{j \epsilon M_i} \mu_{lj}^{m} \left[K\left(i-j\right) \left\| q_i - a_i d_l \right\|^2 + R_{li} \right] \tag{7}$$

In Equation (7), m indicates the existence of m pixels in the image, d indicates the classification of the image into d categories, and μ_{lj} represents the degree to which pixel j belongs to d category. Then the calculated equations for cluster center, membership degree, and offset field can be given as follows:

$$\mu_{lj} = \frac{1}{\sum_{f=1}^{d} \left[\frac{W_{lj} + R_{li}}{W_{fj} + R_{fj}} \right]^{\frac{1}{n-1}}} \tag{8}$$

$$W_{lj} = \sum_{i=1}^{m} K(i-j) \left\| q_i - a_i d_l \right\|^2 \tag{9}$$

$$CC_l = \frac{\sum_{j=1}^{m} \mu_{lj}^{m} q_j \sum_{i=1}^{m} K(i-j) a_i}{\sum_{j=1}^{m} \mu_{lj}^{m} q_j \sum_{i=1}^{m} K(i-j) a_i^2} \tag{10}$$

$$a_i = \frac{\sum_{i=1}^{m} K(i-j) q_j \sum_{j=1}^{m} \mu_{lj}^{m} d_l}{\sum_{i=1}^{m} K(i-j) q_j \sum_{j=1}^{m} \mu_{lj}^{m} d_l^2} \tag{11}$$

Equations (9–11) represent the cluster center, degree of membership, and offset field, respectively. m represents the existence of m pixels in the image, d represents the classification of the image into d categories, and μ_{lj} represents the degree to which pixel j belongs to d category. Based on the possibility of the FCM algorithm MRI segmentation method, the steps were as follows. Step 1: the noise and brain tissue in the brain MRI were removed, and the image was normalized to prepare for the next step. Step 2: the histogram statistics was performed on the three-dimensional MRI data, and the value of 0.707 near the highest value sample was automatically obtained, and the threshold was set to obtain the classification number. Step 3: the initial segmentation was performed through FCM with an uncertain membership relationship to obtain the initial membership degree and cluster center. Step 4:the matrix size of multiple own membership degrees was adjusted. Step 5: the collaborative possibility FCM algorithm was adopted to segment the lesion area to obtain the segmented image. Therefore, the calculation process of the algorithm constructed in this study can be shown in **Figure 1**.

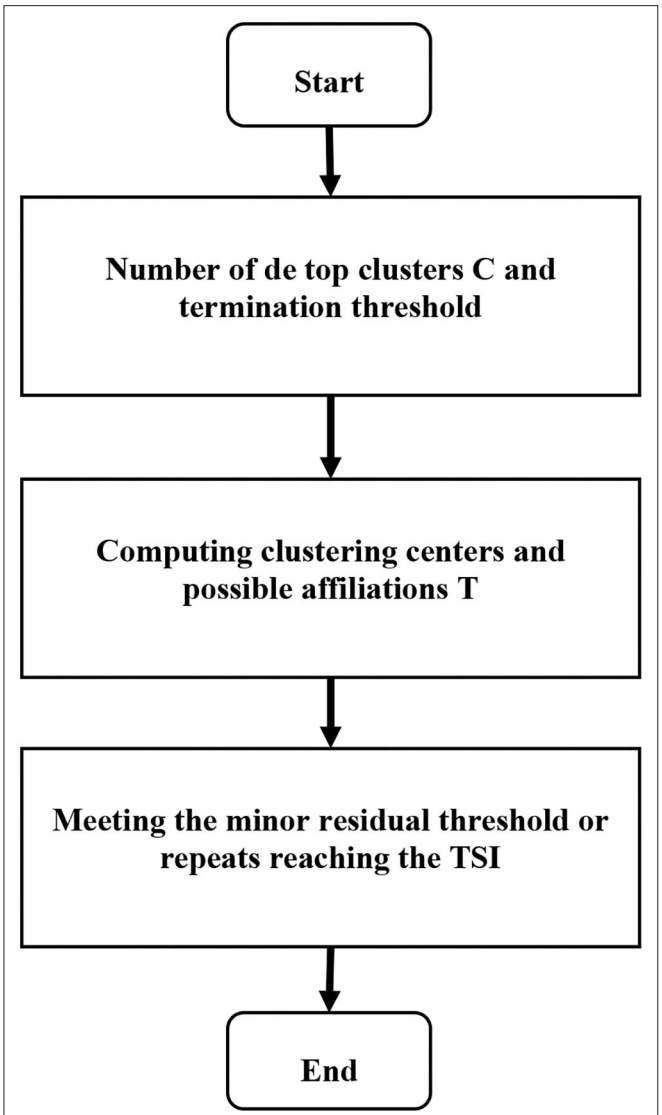


FIGURE 1 | Process of MRI segmentation using the fuzzy C-means clustering (FCM) algorithm.

Assessment of Lacunar Infarction

An MRI examination of the patients' brains was performed. When the sagittal T1WI phase was in a low-signal state, and the axial T2WI and coronary T2 FLAIR were in a high-signal state, it was determined that the diameter was less than 25 mm, and the ischemic infarct lesion is determined. When the diameter was less than 3 mm, it was considered to increase the peripheral blood vessel gap. The numbers of cerebral infarction lesions with a diameter of 3–25 mm were recorded, and the average value was taken as the number of cerebral infarction lesions. The identification and counting were carried out by an experienced radiologist.

Evaluation of White Matter Lesions

An MRI examination of the patients' brains was performed. The T1 weighted was in a low-signal state in the white matter in the bilaterally symmetrical lateral ventricle, whereas the T2 weighted and FLAIR sequence were in a high-signal state with patchy shadows and blurred boundaries. It was determined as the white matter lesion (15). Classification of the white matter lesion was as below: (Level 0: no lesion; Level 1: punctate lesion; Level 2: fusion of lesion; Level 3 lesion merging in large areas).

In this study, segmentation accuracy (SA) and similarity (KI) were used as indicators to evaluate segmentation results. The available equations were given as follows:

$$KI = \frac{2(A \bigcap B)}{2\left(A \bigcap B\right) + (A - B) + (B - A)} \tag{12}$$

$$SA = \sum_{i=1}^{c} \frac{A_i \bigcap B_i}{\sum_{j=1}^{c} B_j} \tag{13}$$

In Equations (12) and (13), A represents the target area pixels obtained by the algorithmic segmentation, and B represents the target area pixels obtained by the standard segmentation.

Statistical Method

In this study, all the experimental data were analyzed by using the SPSS 13.0. The percentage (%) and the mean and SD ($\bar{x} \pm s$) were used to indicate the measured data. The χ^2 test was used to compare the gender, age, and education level of patients in the control, VCIND, VD, and AD groups, and the *t*-test was used to compare the age among four groups. The SPSS Spearman correlation analysis was used to analyze the correlation between the hippocampal volume and age, education level, and ADL, and to understand the relationship of the data. Comparison of data in different groups was analyzed by the one-way ANOVA. If $p < 0.05$, the difference had a statistical significance.

RESULTS

MRI of the Patients

The head MRI of the patients in the control, VD, and AD groups were compared. There was no lacunar infarction, white matter lesions, and hippocampal volume changes for the patients in the control group, as shown in **Figure 2a**. The white matter lesion occurred for the patients in the VD group, as shown in **Figure 2b**. The patients in the AD group had lacunar infarction, white matter lesions, and reduced hippocampal volume, as shown in **Figure 2c**.

Comparison of Algorithm Segmentation Effect

In this study, the standard FCM algorithm was introduced, the bias-corrected FCM (BCFCM) improved based on the global information, and compared the segmentation effect based on the rough FCM (RFCM) algorithm were compared with the algorithm proposed in this study. As shown in **Figure 3**, the improved FCM algorithm used in this study was significantly better than other algorithms in terms of noise removal and detail preservation, and the segmentation effect was the most ideal.

Normal tissues in the brain MRI mainly included cerebral cortex, gray matter, white matter, cerebrospinal fluid, etc.; the

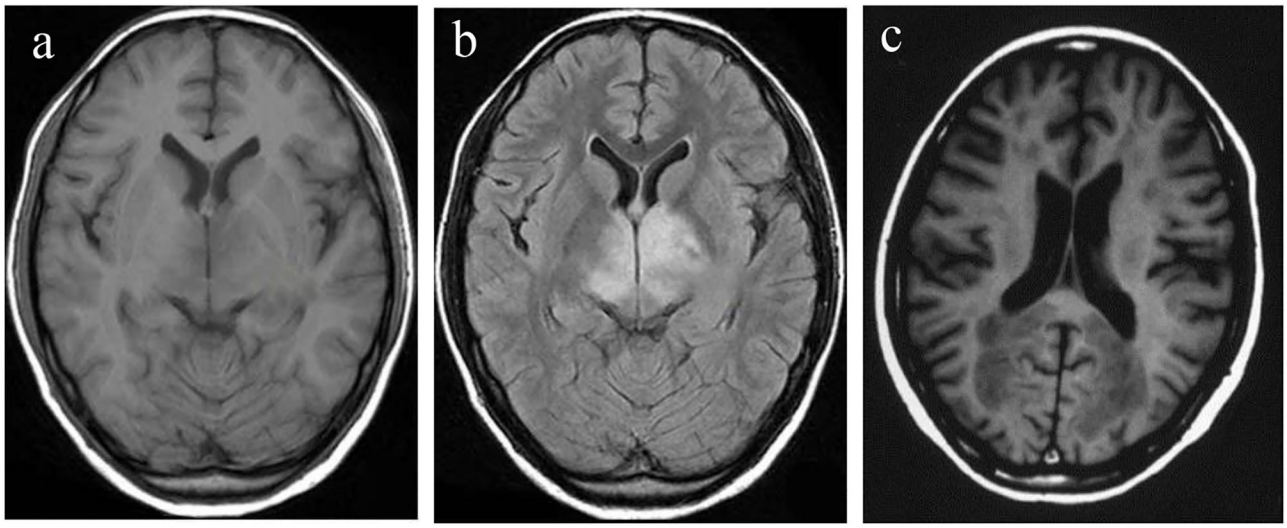


FIGURE 2 | Head MRI of the patients in control, VD, and AD groups. **(a)** Control group; **(b)** VD group; **(c)** AD group. AD, Alzheimer's disease; VD, vascular dementia.

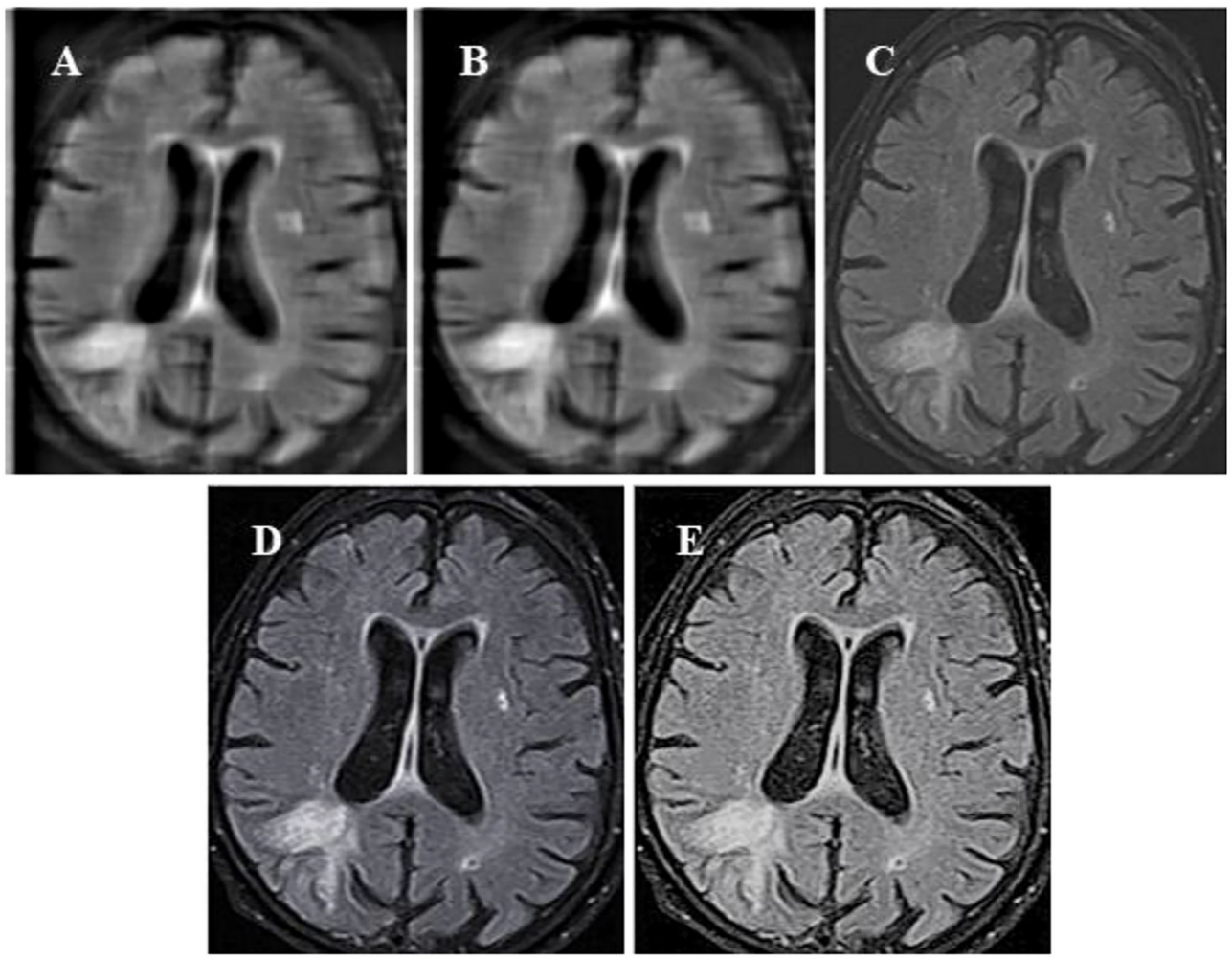


FIGURE 3 | Segmentation results of MRI of brain tumors by the algorithm. **(A)** The original image; **(B)** the standard FCM algorithm; **(C)** the RFCM algorithm; **(D)** the BCFCM algorithm; and **(E)** the improved FCM algorithm. FCM, fuzzy C-means; BCFCM, bias-corrected fuzzy C-means; RFCM, rough fuzzy C-means.

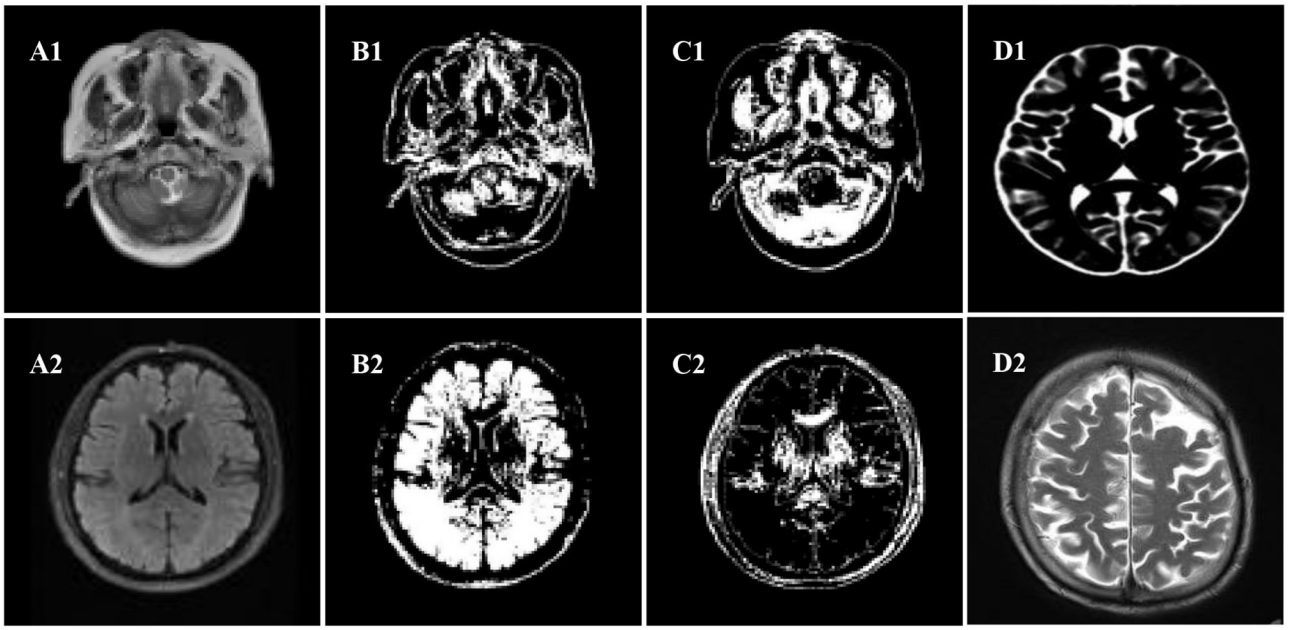


FIGURE 4 | Segmentation effect diagram. **(A1,A2)** Original images, **(B1,B2)** white matter, **(C1,C2)** gray matter, **(D1,D2)** organ.

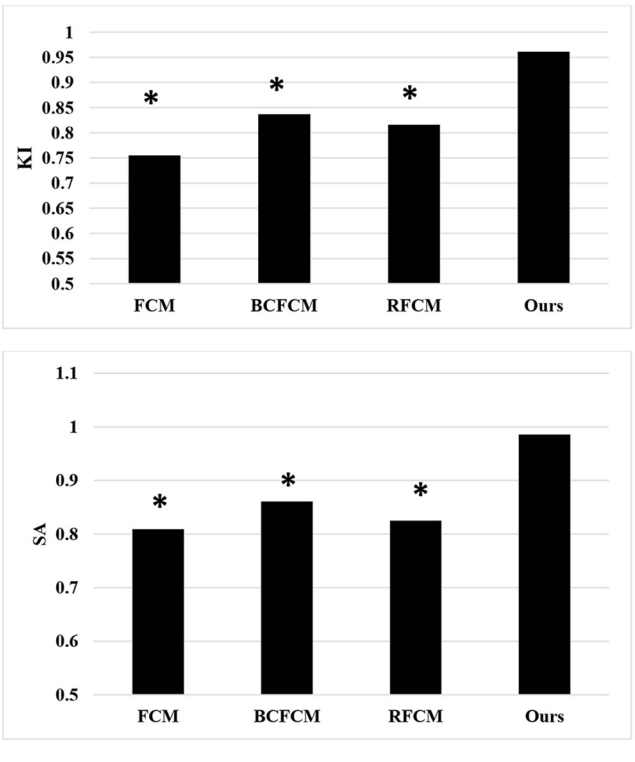


FIGURE 5 | Comparison of segmentation indexes of algorithms. *The difference compared with the algorithm in this study was statistically significant ($p < 0.05$).

diseased tissues mainly included tumors, edema, and necrosis and cystic transformation inside tumors, and the structure of each tissue was newly transmitted and complex. The space occupancy, new transmission, and size of intracranial lesions were uncertain; the gray distribution between different soft tissues or between the diseased tissue and the soft tissue was uneven; and the aliased density distribution and other factors would affect the image segmentation effect. The MRI of the patient's brain was processed by an algorithm, and the comparison of the white matter, gray matter, and organs of the brain was shown in **Figure 4**.

Further comparison of the quantitative data results (**Figure 5**) showed that the KI and SA of the improved FCM algorithm used in this study were 0.961 and 0.986, respectively; those of the FCM algorithm were 0.755 and 0.809, respectively; those of the BCFCM algorithm were 0.837 and 0.861, respectively; and those of the RFCM algorithm were 0.816 and 0.825, respectively. The KI and SA index data of the improved FCM algorithm used in this study were significantly higher than the standard FCM algorithm, BCFCM algorithm, and RFCM algorithm, showing statistically significant differences ($p < 0.05$).

Comparison of Patients' Basic Information

The sex and education level of the patients in the control, VCIND, VD, and AD groups were compared, as shown in **Figure 6**. It was found that there was no significant difference in gender and education level by the χ^2 test ($p > 0.05$), and there was no significant difference in age by the t-test ($p > 0.05$). After analysis

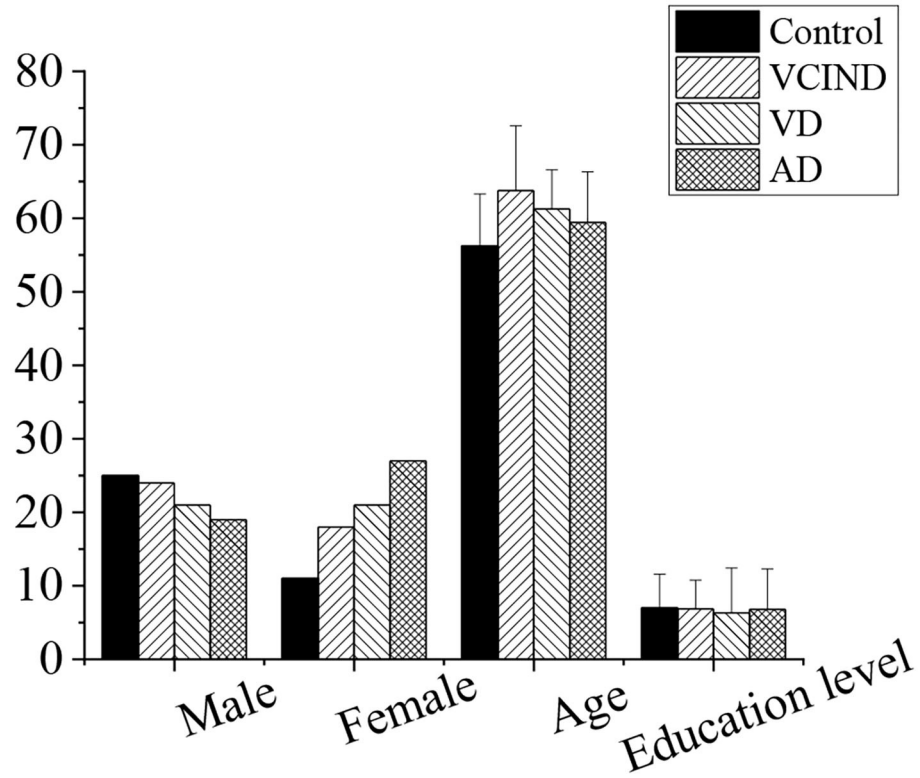


FIGURE 6 | Comparison of patients' basic information in the control, VCIND, VD, and AD groups. There was no statistic difference for the gender, age, and education level ($p > 0.05$). AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia.

of the patient's basic information and data, it was comparable in gender, age, and educational level ($p > 0.05$).

Analysis of Related Risk Factors of Patients in Four Groups

As shown in **Figure 7**, the hypertension, diabetes, smoking history, drinking history, hyperlipidemia, and coronary heart disease of the patients in the control group (46 cases), VCIND group (42 cases), VD group (42 cases), and AD group (46 cases) were analyzed and compared. It was found that the number of patients with associated hypertension in the VCIND group (22 cases, 52.4%), VD group (30 cases, 71.4%), and AD group (17 cases, 37%) was significantly higher than that in the control group (16 cases, 34.8%) with a statistical difference ($^{*}p < 0.05$). The VD group was significantly higher than that in the VCIND group, which had a statistical difference by comparison ($^{\#}p < 0.05$). There were no significant statistical differences among the other related risk factors ($p > 0.05$).

Evaluation of Cognitive Function of the Patients in Four Groups

MOCA, MES, and ADL were evaluated in the control, VCIND, VD, and AD groups, as shown in **Figure 8**.

First, in the MOCA and MES, the scores of patients in the VD group were 15.09 ± 4.33 and 15.11 ± 5.2, respectively, and were higher than those of the patients in the control group, which were MOCA (25.03 ± 5.2) and MES (24.31 ± 3.67) ($^{*}p < 0.05$). The values in the VD group were also lower than those in the VCIND group, which were MOCA (22.17 ± 4.89) and MES (20.08 ± 5.33), with a statistical difference ($^{\#}p < 0.05$).

Second, in the ADL, the values in the VCIND group (23.55 ± 6.12) and the VD group (28.56 ± 3.1) were higher than that in the control group (19.17 ± 3.67), and it had a statistical difference ($^{\Delta}p < 0.05$). There was no significant difference compared between the VCIND group and the VD group ($p > 0.05$).

Third, there was no statistical difference for MOCA, MES, and ADL in the AD group from the control, VCIND, and VD groups ($p > 0.05$).

Analysis of Distribution of Lacunar Cerebral Infarction and Number of Lesions of Patients in Four Groups

The distribution of lacunar infarction and the number of lesions of patients in the control group (46 cases), VCIND group (42 cases), VD group (42 cases), and AD group (46 cases) were analyzed, as shown in **Figure 9**. The ratios of the lacunar cerebral infarction distribution on both sides of the brain in the VCIND group (22 cases, 52.4%), VD group (13 cases, 31%), and AD group (12 cases, 26.1%) were significantly higher than that in

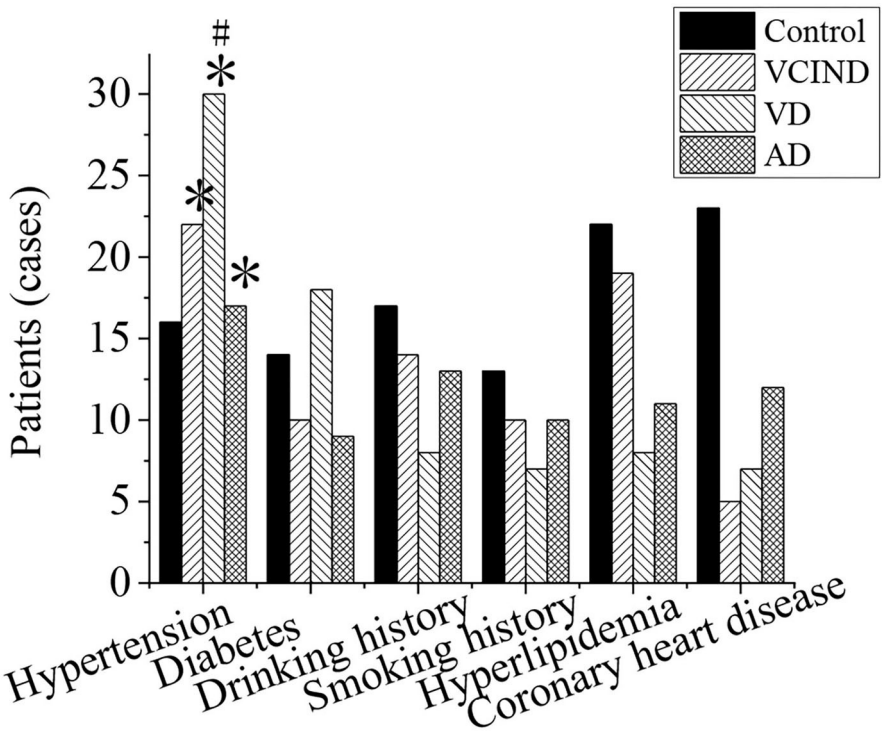


FIGURE 7 | Comparison of related risk factors of patients in the control, VCIND, VD, and AD groups. $^{*}p < 0.05$ if compared with the control group, whereas $^{\#}p < 0.05$ if compared with the VCIND group. AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia.

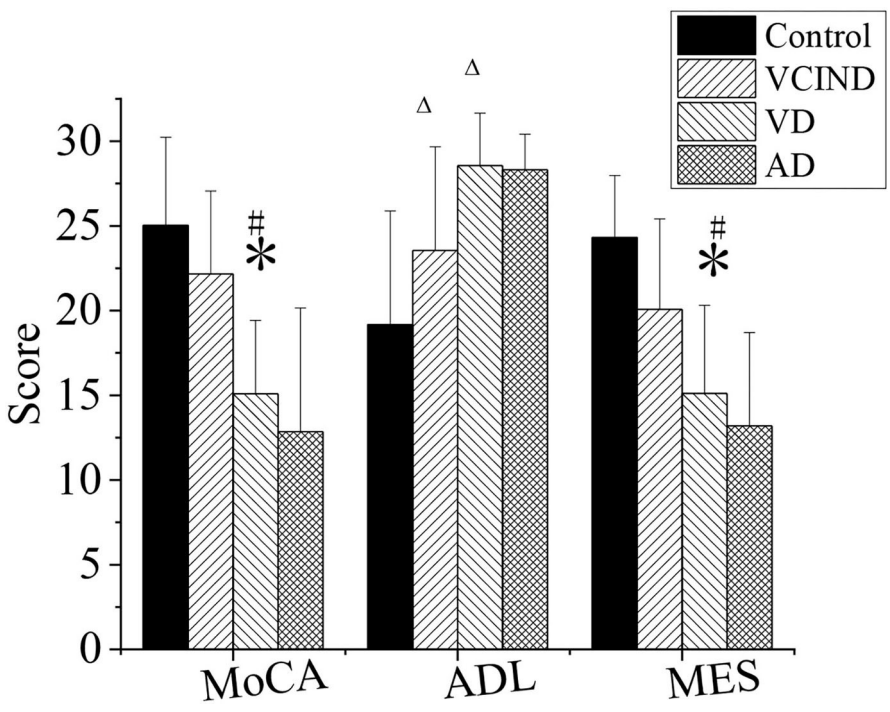


FIGURE 8 | Comparison of cognitive functions of patients in the control, VCIND, VD, and AD groups. AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia. $^{*}p < 0.05$ if compared with MMSE, MOCA, and MES in the control group, whereas $^{\#}p < 0.05$ if compared with the VCIND group. $^{\Delta}p < 0.05$ if compared with ADL in the control group.

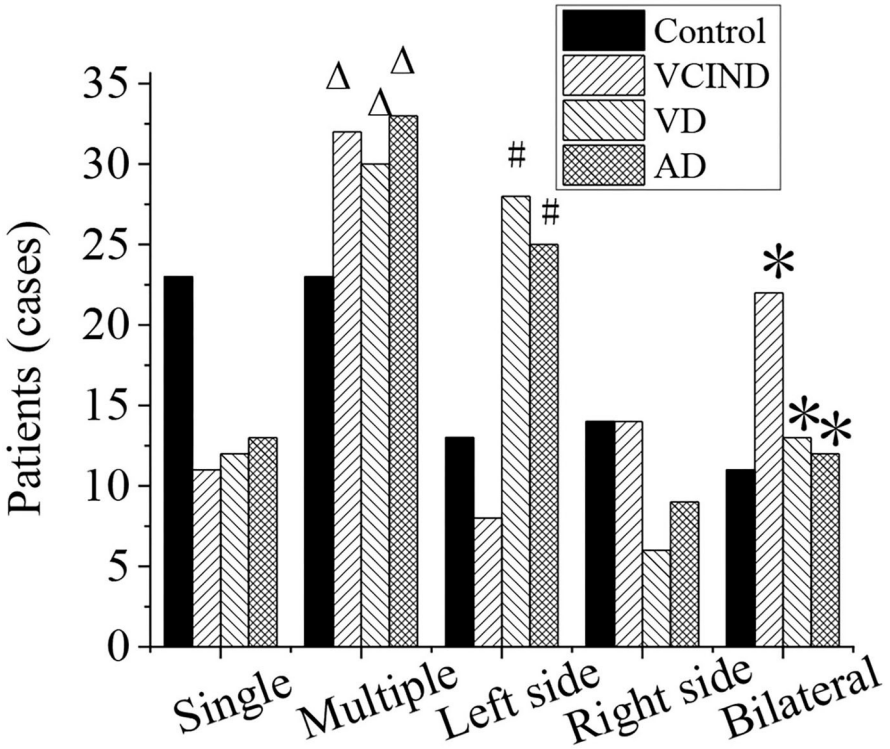


FIGURE 9 | Analysis of distribution of lacunar cerebral infarction and number of lesions of patients in the control, VCIND, VD, and AD groups. AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia. $^{*}p < 0.05$ if compared with distribution of lacunar cerebral infarction in the control group, whereas $^{\#}p < 0.05$ if compared with the VCIND group. $^{\Delta}p < 0.05$ if compared with number of lesions in the control group.

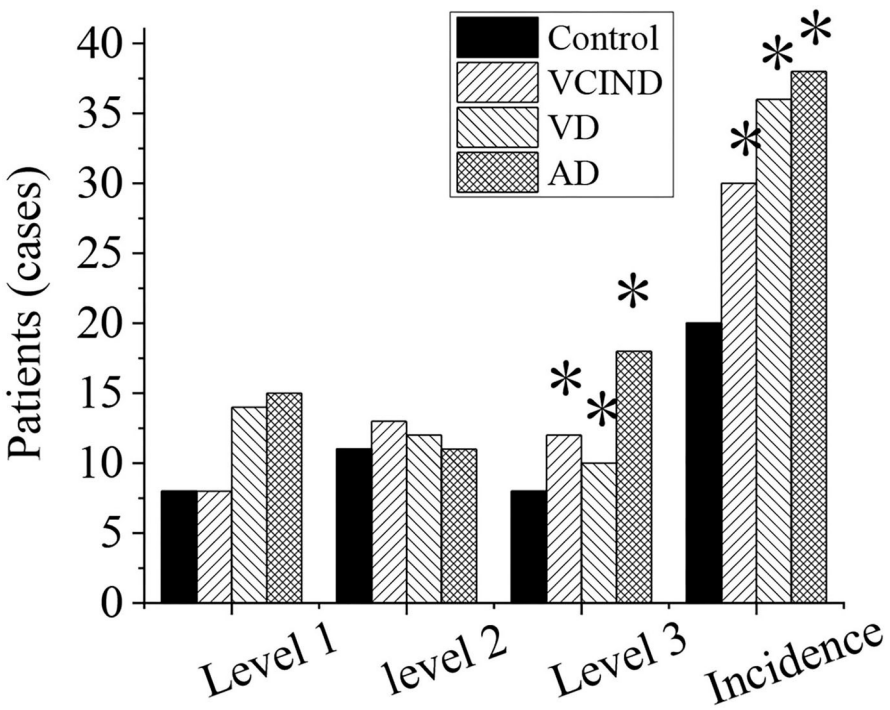


FIGURE 10 | Analysis of white matter lesions and incidence rate of patients in the control, VCIND, VD, and AD groups. AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia. $^{*}p < 0.05$ if compared with white matter lesions and incidence rate in the control group.

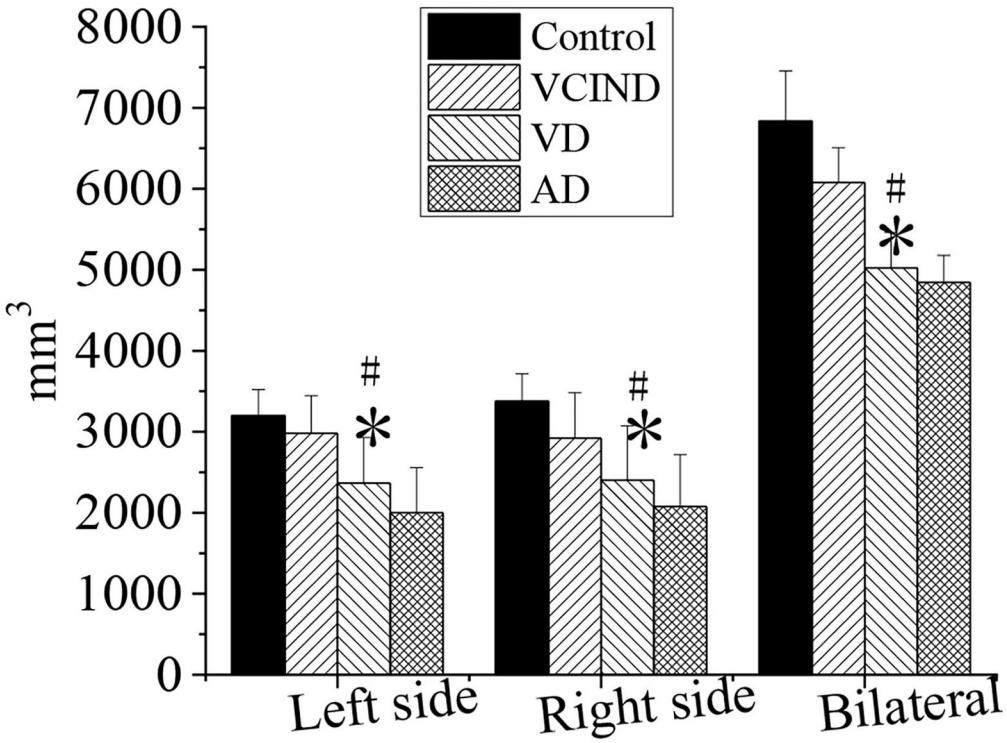


FIGURE 11 | Comparison of hippocampal volumes of patients in the control, VCIND, VD, and AD groups. AD, Alzheimer's disease; VCIND, vascular cognitive impairment non-dementia; VD, vascular dementia. $^{*}p < 0.05$ if compared with the left, right, and bilateral hippocampal volumes in the control group, whereas $^{\#}p < 0.05$ if compared with the VCIND group.

the control group (11 cases, 23.9%), which had a statistical difference ($^{*}p < 0.05$). The proportions of the lacunar cerebral infarction on the left side of the brain in the VD group (28 cases, 66.7%) and AD group (25 cases, 54.3%) were higher than that in the VCIND group (8 cases, 19%), which had a statistical difference ($^{\#}p < 0.05$).

The numbers of lacunar infarction lesions in the VCIND group (32 cases, 76.2%), VD group (30 cases, 71.4%), and AD group (33 cases, 71.7%) were significantly higher than that in the control group (23 cases, 50%), and it had a statistical difference ($^{\Delta}p < 0.05$).

Analysis of White Matter Lesions and Incidence Rate of the Patients in Four Groups

The white matter lesions and incidence rates of patients in the control group (46 cases), VCIND group (42 cases), VD group (42 cases), and AD group (46 cases) were compared and analyzed, as shown in **Figure 10**. Compared with the control group (8 cases, 17.4%), the proportions of 3-grade lesions in the VCIND group (12 cases, 28.6%), VD group (10 cases, 23.8%), and AD group (18 cases, 42.9%) were increased, which had a statistical difference ($^{*}p < 0.05$). Compared with the control group (20 cases, 43.5%), the incidence rates of lesions in the VCIND group (30 cases, 71.4%), VD group (36 cases, 85.7%), and AD group (38 cases, 82.6%) were increased, which had statistical differences ($^{*}p < 0.05$).

Comparison of Hippocampal Volumes of Patients in Four Groups

The hippocampal volumes (left, right, and bilateral hippocampal volumes) of patients in the control, VCIND, VD, and AD groups were compared and analyzed, as shown in **Figure 11**. It compared the left (2,367 ± 563), right (2,401 ± 672), and bilateral hippocampal volume (5,023 ± 437) of patients in the VD group with the left (3,201 ± 321), right (3,378 ± 337), and bilateral hippocampal volume (6,834 ± 621) of patients in the control group. The results revealed that there was a statistical difference ($^{*}p < 0.05$). By comparing the left (2,982 ± 463), right (2,921 ± 564), and bilateral hippocampus volume (6,079 ± 429) of the patients in the VCIND group, the left, right, and bilateral hippocampus volume of the patients in the VD group had a statistical difference ($^{\#}p < 0.05$).

Relationship Between Hippocampal Volume and Age and Education Level

As shown in **Figure 12**, Pearson's correlation analysis was taken for the relationship between the hippocampus volume and age, and it was found that the age and hippocampus volume were negatively correlated ($r = -0.919$, $p < 0.01$). With the increase of age, hippocampus volume tended to decrease.

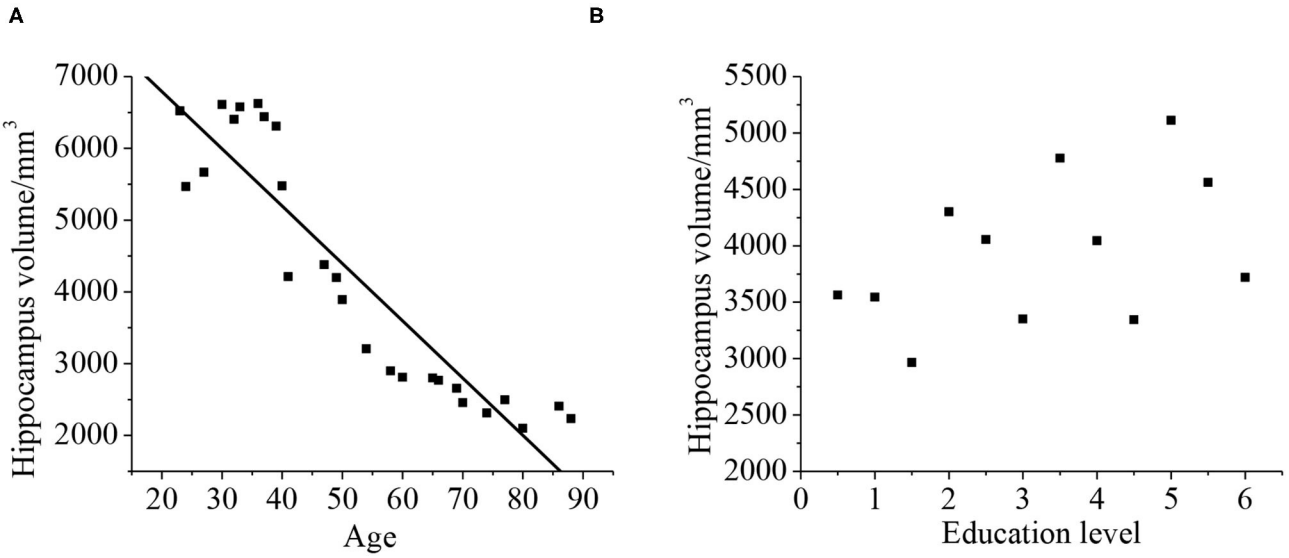


FIGURE 12 | Correlation between the hippocampal volume, age **(A)**, and education level **(B)**. The correlation existed if $p < 0.01$.

The hippocampal volume has no phase relationship with the education level ($p > 0.01$).

Relationship Between Hippocampal Volume and the Neuropsychological Assessment ADL

Pearson's correlation analysis was taken for the relationship between the hippocampal volume and neuropsychological assessment. It was found that the neuropsychological assessment ADL was negatively correlated with hippocampal volume ($r = -0.872$, $p < 0.01$). With less hippocampal volume of the patient, both the cognitive ability and daily living ability of the patient decrease, as shown in **Figure 13**.

DISCUSSION

Among fuzzy clustering, the most commonly used is FCM. However, FCM also has some shortcomings. The sum of the membership degrees of the same sample belonging to all classes is 1, which makes it sensitive to noise and outliers; and it is sensitive to the initial cluster center (18). In order to overcome these shortcomings, some scholars have proposed possible C-means clustering, weighted FCM clustering, and FCM based on the possibility of membership uncertainty. These clustering algorithms are based on the clustering of all special diagnoses in the dataset, and the weight of each special diagnosis is different (19, 20). In order to make the clustering results more accurate, the possibility of collaboration FCM is a method of clustering using the collaborative relationship between different features (12, 21). This method can be combined with other clustering algorithms to improve the clustering effect. The fuzzy partition matrix is obtained by using the synergy coefficient between the different features themselves (22). The detection by the MRI based on artificial intelligence technology will play an important role in the early diagnosis and treatment of diseases related to vascular cognitive impairment. Artificial intelligence can effectively complete the collection of relevant data information by simulating human thinking instead of an artificial collection. In this study, the relevant case data of patients in the Department of Neurology and Physical Examination who entered Changzhi People's Hospital from October 1, 2018, to February 1, 2020 were selected. Combined with the MRI data based on artificial intelligence technology, the risk factors of patients in the control, VCIND, VD, and AD groups were analyzed. It was found that the numbers of patients with associated hypertension in the VCIND group (52.4%), VD group (71.4%), and AD group (37%) are significantly higher than that in the control group (34.8%), which has a statistical difference ($^{*}p < 0.05$). That in the VD group was significantly higher than that in the VCIND group, which had a statistical difference by comparison ($^{\#}p < 0.05$). There were no significant statistical differences among the other related risk factors ($p > 0.05$). It is consistent with the previous results, suggesting that it plays a key role in controlling hypertension in the prevention and control of cognitive impairment (23). If the blood pressure is under high pressure for a long time without any control, the elasticity of small cerebral blood vessels will be reduced accordingly, and the blood circulation will be poor. It will lead to hypoxia in the brain, lacunar cerebral infarction, and white matter lesions and develop into a VD. Some related studies show that diabetes, hyperlipidemia, and coronary heart disease are related to cognitive impairment (24), which is inconsistent with the results of this study. No correlation between gender, age, and education level to the cognitive impairment was found in this study.

In this study, a comprehensive evaluation of the cognitive ability is taken for all patients. It is found the scores of MOCA

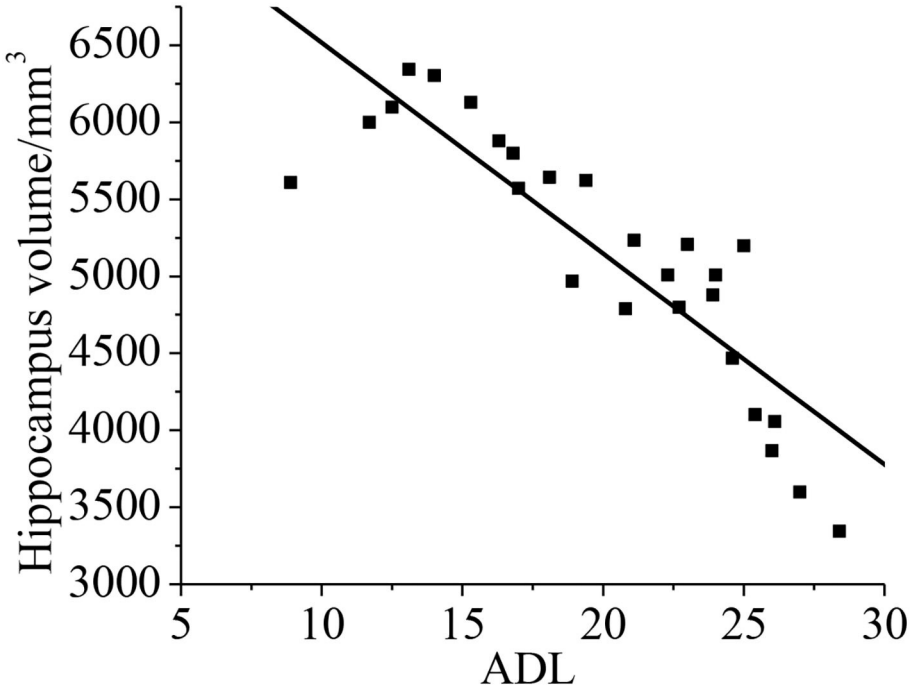


FIGURE 13 | Correlation between the hippocampal volume ADL. The correlation existed if $p < 0.01$.

and MES in the VD group are 15.09 ± 4.33 and 15.11 ± 5.2, respectively. They are lower than that in the control group, and there is a statistical significance ($p < 0.05$). It indicates that compared with the control group, the cognitive level in the VD group decreases greatly, whereas the cognitive level in the VCIND decreases slightly. Such results are consistent with the study of Jadavji et al. (25). In the ADL, the scores in the VCIND group (23.55 ± 6.12) and the VD group (28.56 ± 3.1) are higher than that in the control group (19.17 ± 3.67), which has a statistical difference ($p < 0.05$). It means that compared with the control group, the daily living ability of patients in the VCIND group and the VD group are significantly decreased. The hippocampal volume is negatively correlated with ADL ($r = -0.872$, $p < 0.01$), indicating that the more serious the hippocampus atrophy, the worse the daily living ability of the patient. Such results are consistent with the results of Wierenga et al. (26). The hippocampal volume is negatively correlated with age ($r = -0.919$, $p < 0.01$), indicating that the hippocampus-related neurons may be lost as age increases, causing their volume to shrink.

The distribution of lacunar infarction and the number of lesions of the patients in the four groups were analyzed. It was found that the proportions of lacunar infarction in the brain (52.4, 31, and 26.1%) and the number of multiple lacunar infarction lesions (76.2, 71.4, and 71.7%) in VCIND, VD, and AD groups were significantly higher than those in the control group (23.9% and 50%). It has a statistical difference ($p < 0.05$). With the increase of the infarction, the incidence rate of the cognitive impairment also increases, and the degree of dementia increases. The bilateral and multiple lesions have more serious damage to the executive function of the patient. It may be due to the multiple arterial ischemias, and more neural network structures are also destroyed.

Compared with the control group, the proportions of 3-grade lesions (28.6, 23.8, and 42.9%) and the incidence rate of lesions (71.4, 85.7, and 82.6%) in the VCIND, VD, and AD groups increased, showing statistics differences. When the white matter of the brain suffers some lesion, the white matter fibers will be destroyed, and the cognitive function will be impaired. It is consistent with multiple studies that white matter lesions are related to cognitive impairment in patients with dementia and non-dementia.

CONCLUSION

In this study, the lacunar cerebral infarction, white matter lesions, and hippocampal volume of the patient with the neurological cerebrovascular disease were studied, and the correlation with cognitive impairment of the patient was analyzed through MRI based on artificial intelligence technology combined with the neuropsychological assessment. However, the shortcoming of this study was that the patient samples involved were fewer, which may cause deviations in the results. Therefore, the results needed to be verified further by the multicenter experimental data with more samples. In short, the improved FCM algorithm had a higher segmentation effect and SA on MRI of neurovascular cerebrovascular diseases. The distribution, number, white matter lesions, and hippocampal volume of lacunar cerebral infarction

were related to the cognitive impairment of patients with cerebrovascular diseases.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Changzhi Medical College. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LZ wrote the manuscript. YL, LB, and QL collected and analyzed general data of patients. XZ and BZ were responsible for the follow-up index analysis. BZ helped with statistical analysis. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Scientific Research Project of Shanxi Provincial Health Commission in 2021 (No. 2021012) and the Key Special Project of Four Batches of Science and Technology Innovation Plan of Shanxi Provincial Health Commission (No. 2020XM38).

REFERENCES

1. Smith EE, Beaudin AE. New insights into cerebral small vessel disease and vascular cognitive impairment from MRI. *Curr Opin Neurol.* (2018) 31:36–43. doi: 10.1097/WCO.0000000000000513
2. Ferro DA, van Veluw SJ, Koek HL, Exalto LG, Biessels GJ, Utrecht Vascular Cognitive Impairment (VCI) Study Group. Cortical cerebral microinfarcts on 3 tesla MRI in patients with vascular cognitive. *Impairment J Alzheimer's Dis.* (2017) 60:1443–50. doi: 10.3233/JAD-170481
3. Wang Y, Tu D, Du J, Han X, Sun Y, Xu Q, et al. Classification of subcortical vascular cognitive impairment using single MRI sequence and deep learning convolutional neural networks. *Front Neurosci.* (2019) 13:627. doi: 10.3389/fnins.2019.00627
4. Wang Y, Cao W, Sun Y, Chen X, Ding W, Xu Q, et al. White matter integrity in subcortical vascular cognitive impairment: a multimodal structural MRI Study. *Curr Alzheimer Res.* (2017) 14:991–9. doi: 10.2174/1567205014666170329113707
5. Pasi M, Sugita L, Xiong L, Charidimou A, Boulouis G, Pongpitakmetha T, et al. Abstract TMP111: prediction of cognitive impairment after intracerebral hemorrhage using MRI small vessel disease score. *Stroke.* (2019) 50:111. doi: 10.1161/str.50.suppl_1.TMP111
6. Bouvy WH, Kuijf HJ, Zwanenburg JJ, Koek HL, Kappelle LJ, Luijten PR, et al. Abnormalities of cerebral deep medullary veins on 7 tesla MRI in amnestic mild cognitive impairment and early Alzheimer's Disease: a pilot study. *J Alzheimer's Dis.* (2017) 57:705–10. doi: 10.3233/JAD-160952
7. Heiss WD, Rosenberg GA, Thiel A, Berlot R, de Reuck J. Neuroimaging in vascular cognitive impairment: a state-of-the-art review. *BMC Med.* (2016) 14:174. doi: 10.1186/s12916-016-0725-0
8. van Veluw SJ, Heringa SM, Kuijf HJ, Koek HL, Luijten PR, Biessels GJ. Cerebral cortical microinfarcts at 7 Tesla MRI in patients with early Alzheimer's disease. *J Alzheimer's Dis.* (2014) 39:163–7. doi: 10.3233/JAD-131040
9. Zuloaga KL, Zhang W, Yeiser LA, Stewart B, Kukino A, Nie X, et al. Neurobehavioral and imaging correlates of hippocampal atrophy in a mouse model of vascular cognitive impairment. *Transl Stroke Res.* (2015) 6:390–8. doi: 10.1007/s12975-015-0412-z
10. Wardlaw JM, Valdés Hernández MC, Muñoz-Maniega S. What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. *J Am Heart Assoc.* (2015) 4:e001140. doi: 10.1161/JAHA.114.001140
11. Bouvy WH, van Veluw SJ, Kuijf HJ, Zwanenburg JJ, Kappelle JL, Luijten PR, et al. Microbleeds colocalize with enlarged juxtacortical perivascular spaces in amnestic mild cognitive impairment and early Alzheimer's disease: a 7 Tesla MRI study. *J Cerebral Blood Flow Metab.* (2020) 40:739–46. doi: 10.1177/0271678X19838087
12. Ahmed HA, Ishrat T, Pillai B, Bunting KM, Patel A, Vazdarjanova A, et al. Role of angiotensin system modulation on progression of cognitive impairment and brain MRI changes in aged hypertensive animals—a randomized double-blind pre-clinical study. *Behav Brain Res.* (2018) 346:29–40. doi: 10.1016/j.bbr.2017.12.007
13. Wisse LE, Reijmer YD, ter Telgte A, Kuijf HJ, Leemans A, Luijten PR, et al. Hippocampal disconnection in early Alzheimer's disease: a 7 tesla MRI study. *J Alzheimer's Dis.* (2015) 45:1247–56. doi: 10.3233/JAD-142994
14. Zhou H, Yang J, Xie P, Dong Y, You Y, Liu J. Cerebral microbleeds, cognitive impairment, and MRI in patients with diabetes mellitus. *Clin Chim Acta.* (2017) 470:14–9. doi: 10.1016/j.cca.2017.04.019
15. Carnevale L, Lembo G. Innovative MRI techniques in neuroimaging approaches for cerebrovascular diseases and vascular cognitive impairment. *Int J Mol Sci.* (2019) 20:2656. doi: 10.3390/ijms20112656
16. Gorelick PB, Counts SE, Nyenhuis D. Vascular cognitive impairment and dementia. *Biochim Biophys Acta.* (2016) 1862:860–8. doi: 10.1016/j.bbadis.2015.12.015
17. Rosenberg GA. Matrix metalloproteinase-mediated neuroinflammation in vascular cognitive impairment of the Binswanger type. *Cell Mol Neurobiol.* (2016) 36:195–202. doi: 10.1007/s10571-015-0277-4
18. Sun Y, Cao W, Ding W, Wang Y, Han X, Zhou Y, et al. Cerebral blood flow alterations as assessed by 3D ASL in cognitive impairment in patients with subcortical vascular cognitive impairment: a marker for disease severity. *Front Aging Neurosci.* (2016) 8:211. doi: 10.3389/fnagi.2016.00211
19. de Bruijn RF, Akoudad S, Cremers LG, Hofman A, Niessen WJ, van der Lugt A, et al. Determinants, MRI correlates, and prognosis of mild cognitive impairment: the Rotterdam Study. *J Alzheimer's Dis.* (2014) 42:S239–49. doi: 10.3233/JAD-132558
20. López-Gil X, Amat-Roldan I, Tudela R, Castañé A, Prats-Galino A, Planas AM, et al. DWI and complex brain network analysis predicts vascular cognitive impairment in spontaneous hypertensive rats undergoing executive function tests. *Front Aging Neurosci.* (2014) 6:167. doi: 10.3389/fnagi.2014.00167
21. Kalaria RN. Neuropathological diagnosis of vascular cognitive impairment and vascular dementia with implications for Alzheimer's disease. *Acta Neuropathol.* (2016) 131:659–85. doi: 10.1007/s00401-016-1571-z
22. Ai Q, Pu YH, Sy C, Liu LP, Gao PY. Impact of regional white matter lesions on cognitive function in subcortical vascular cognitive impairment. *Neurol Res.* (2014) 36:434–43. doi: 10.1179/1743132814Y.0000000354
23. Malojcic B, Giannakopoulos P, Sorond FA, Azevedo E, Diomedi M, Oblak JP, et al. Ultrasound and dynamic functional imaging in vascular cognitive impairment and Alzheimer's disease. *BMC Med.* (2017) 15:27. doi: 10.1186/s12916-017-0799-3

24. Jang YK, Kwon H, Kim YJ, Jung NY, Lee JS, Lee J, et al. Early-vs late-onset subcortical vascular cognitive impairment. *Neurology*. (2016) 86:527–34. doi: 10.1212/WNL.0000000000002357
25. Jadavji NM, Farr TD, Lips J, Khalil AA, Boehm-Sturm P, Foddis M, et al. Elevated levels of plasma homocysteine, deficiencies in dietary folic acid and uracil–DNA glycosylase impair learning in a mouse model of vascular cognitive impairment. *Behav Brain Res*. (2015) 283:215–26. doi: 10.1016/j.bbr.2015.01.040
26. Wierenga CE, Hays CC, Zlatar ZZ. Cerebral blood flow measured by arterial spin labeling MRI as a preclinical marker of Alzheimer's disease. *J Alzheimer's Dis*. (2014) 42:S411–9. doi: 10.3233/JAD-141467

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Li, Bian, Luo, Zhang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fpubh.2022.861062

A Decision Support System for Diagnosing Diabetes Using Deep Neural Network

*Osama Rabie[1], Daniyal Alghazzawi[1], Junaid Asghar[2], Furqan Khan Saddozai[3] and Muhammad Zubair Asghar[3]**

[1] Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, [2] Faculty of Pharmacy, Gomal University, Dera Ismail Khan, Pakistan, [3] Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan, Pakistan

Background and Objective: According to the WHO, diabetes mellitus is a long-term condition marked by high blood sugar levels. The consequences might be far-reaching. According to current increases in mortality, diabetes has risen to number 10 among the leading causes of mortality worldwide. When used to predict diabetes using unbalanced datasets from testing, machine learning (ML) classifiers and established approaches for encoding categorical data have exhibited a broad variety of surprising outcomes. Early studies also made use of an artificial neural network to extract features without obtaining a grasp of the sequence information.

Methods: This study offers a deep learning-based decision support system (DSS), utilizing bidirectional long/short-term memory (BiLSTM), to accurately predict diabetic illness from patient data. In order to predict diabetes, the BiLSTM hybrid model was used after balancing the data set.

Results: Unlike earlier studies, this proposed model's trial findings were promising, with an accuracy of 93.07%, 93% precision, 92% recall, and a 92% F1-score.

Conclusions: Using a BILSTM model for classification outperforms current approaches in the diabetes detection domain.

Keywords: disease diagnoses, deep learning, diabetes prediction, decision support system, disease diagnosis

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Rakesh Kumar Mahendran, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, India
Waqas Haider Khan Bangyal, University of Gujrat, Pakistan
Praveen Kumar, VIT University, India

***Correspondence:**
Muhammad Zubair Asghar
zubair@gu.edu.pk

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *24 January 2022*
Accepted: *07 February 2022*
Published: *17 March 2022*

Citation:
Rabie O, Alghazzawi D, Asghar J, Saddozai FK and Asghar MZ (2022) A Decision Support System for Diagnosing Diabetes Using Deep Neural Network. Front. Public Health 10:861062. doi: 10.3389/fpubh.2022.861062

INTRODUCTION

With the emergence of AI, data mining applications have become more prevalent in several fields, such as business, education, and healthcare. Healthcare decision support systems are a hot research issue because they allow the finding of exciting patterns and useful data from enormous quantities of healthcare records. Decision support systems might help human medical specialists diagnose illnesses faster by transforming data sources into relevant insights (1). In the fast expanding discipline of data mining known as "deep learning" (DL), a complex mix of feature encoding approaches is used in order to understand from prior data to produce correct estimates (2). There are a number of uses for it, including sentiment classification (3), smart agriculture (4), and more. Recently, neural network models have shown a remarkable capability for content prediction and classification. As shown in Shickel et al. (5) and Miotto et al. (6), deep learning algorithms have figured prominently in healthcare for knowledge discovery and disease diagnosis, such as cardiac diseases, psychiatric disorders, and diabetes disorders, using health data.

A Need for Diabetes Disease Prediction

Diabetes affects more than 34 million people in the United States, accounting for approximately 11% of the population. Diabetes is diagnosed in the United States at a rate of 17 cases per second. Each year, around 1.5 million Americans are diagnosed with diabetes (7). There are several biological indicators and risk factors that must be considered in order to get a definitive diagnosis of diabetes. Some of these indications and variables include age, gender identity, hypertension, and cholesterol levels, among others.

Following China and India, Pakistan has the third highest diabetes prevalence (8). The International Diabetes Federation (IDF) estimates that about 33 million Pakistanis have diabetes. Diabetes diagnoses are crucial since a patient's life is at stake. Diabetics must emphasize early diagnosis and treatment. Numerous complications of diabetes, including nephropathy, retinopathy, neuropathy, cardiovascular disease, stroke, and death, can be delayed or avoided with proper management of elevated blood sugar levels. More effective supervised learning algorithms for disease prediction can significantly reduce these medical errors. The healthcare sector has been supplied with a variety of supervised learning methods by researchers (9). In the health sector, data scientists are encouraged to build useful applications that may help healthcare experts diagnose and manage diabetes illness (10). To effectively predict diabetes illness, it is vital that state-of-the-art deep learning (DL) techniques be researched and applied to health-related patient data.

The Study's Goals

In the past several years, a number of scientists have investigated the possibility of using health data to predict diabetes using computational approaches like machine learning (ML) (9, 10). This research's major goal was to discover ways to detect diabetes before symptoms appear. Additionally, they were constrained by traditional encoders that did not adequately handle the relationship among the disease dataset's predictors. Consequently, the proposed BILSTM-based framework in this research accurately diagnoses diabetes.

Baseline Investigation

Butt et al. developed an ML-based prediction approach for diabetic disease (9). ML was used to better predict diabetes using random forest (RF), support vector machine (SVM), and other classifiers. However, ML classifiers use a traditional encoding strategy that fails to account for the predictors' underlying links. As a result, typical machine learning algorithms are often inefficient at accurately predicting diabetes risk from medical information. Because this study effort had certain limitations (9), we proposed an updated DL model called BiLSTM, which has previously been effectively employed in several fields such as DDoS attack prediction, behavior recognition, and others (2, 11). To predict diabetes, we created the BiLSTM model.

Butt et al. developed an ML-based prediction approach for diabetic disease (9). ML was used to better predict diabetes using random forest (RF), support vector machine (SVM), and other classifiers. However, ML classifiers use a traditional encoding strategy that fails to account for the predictors' underlying links. As a result, typical machine learning algorithms are often inefficient at accurately predicting diabetes risk from medical information. Because this study effort had certain limitations (9), we proposed an updated DL model called BiLSTM, which has previously been effectively employed in several fields such as DDoS attack prediction, behavior recognition, and others (2, 11). To predict diabetes, we created the BiLSTM model.

Problem Statement

The use of traditional feature sets followed by an ML classifier makes it difficult to accurately predict diabetes from patient data (9, 10). Furthermore, the lack of relevant context makes DL models for diabetic illness prediction less effective. Predicting diabetes from patient data is treated as a binary-label prediction problem to handle the aforementioned difficulties. Diabetes is predicted from the supplied illness dataset. There are two classes of data in the dataset: D1 (yes, you have diabetes) and D2 (no, you do not have diabetes). The neural network uses these two classes to predict whether or not someone has diabetes. Using a deep neural network, we want to develop an automated system that can learn from training data and predict the presence or absence of diabetic illness using context information in the healthcare sector.

Research Objectives

We intend to fulfill the following research goals in order to be able to perform an effective diabetic diagnosis.

RO1: To use the BiLSTM deep learning model to make predictions about diabetes based on patient illness data.
RO2: Comparison of the BiLSTM model for diabetes prediction with classical machine learning and deep learning.
RO3. Comparison of proposed method's effectiveness to baseline research for predicting diabetic patients.

Research Contributions

The following are key contributions made by this work:

1. A deep learning (BILSTM) system is being developed to diagnose diabetic disease.
2. The proposed deep learning model for diabetes diagnosis outperforms existing classical machine learning models in terms of prediction performance.
3. To make a diabetes prognosis, two decision classes are employed.
4. Comparing the proposed strategy to existing deep learning and benchmark studies.
5. Using the proposed strategy significantly improves the model's accuracy in predicting diabetes.

The remainder of the study is organized in the following order: Section Related Work gives a review of the current literature, and Section Proposed Methodology discusses the recommended technique. Results and discussion are presented in Section Experimental Results and Discussion, and future applicability of the proposed approach is discussed in Section Conclusions and Future Work.

RELATED WORK

Past studies on diabetes disease prediction are summarized in this section.

Qawqzeh et al. (12) suggested a photoplethysmogram-based regression model for diabetes diagnosis. The framework was validated and evaluated using input from 450 participants and 130 pieces of information. Their suggested approach properly identified 550 non-diabetics with 92% accuracy. But the suggested approach is not compared to current methods. Automated categorization of diabetes using a machine learning technique was given by (13). They employed a SVM classifier using hyperglycemia samples from the UCI Machine Archive. It outperformed Naive Bayes, decision trees, and neural nets. While a contrast of latest systems is provided, there is no discussion of parameter estimation. An SVM-based classifier was employed by Gupta et al. (10) to identify diabetes. They made use of PIMA Indian Diabetes as a resource. Additional methods for improving predictive performance included variable selection and k-fold cross-validation. During the tests, the support vector machine did better than the naive Bayes model. In contrast, there is a lack of current comparability and consistency. Choubey et al. (14) compared numerous diabetes classification systems. The UCI Machine Learning Repository's PIMA Indian collection was integrated with an indigenous hyperglycemia collection. The researchers used SVM, KNN, and NB to identify insulin-dependent individuals from pooled datasets. PCA and LDA feature engineering approaches have been found to improve classification system performance and remove redundant features. Butt et al. (9) looked at utilizing machine learning to diagnose and forecast diabetes. It also showcases an Internet-of-things diabetic tracking device for both normal and sick people. Diabetes was classified using three classification methods: randomized forest, multilayer perceptron, and regression models (LR). They employed SVMs, MA, and linear regression to predict outcomes (LR). The study used the PIMA Indian Diabetes dataset. MLP outperforms similar learners with an accuracy of 86%, whereas LSTM outperforms others with an average of 87%. Zhou et al. (15) proposed a DTP model for glycemic control diagnosis. Each of the data sets contained over 1,000 entries. Smaller epochs in the training step ensure that the technique works rapidly on any smartphone. The findings confirm the effectiveness of the suggested model. Mujumdar and Vaidehi (16) suggested a diabetes prediction model for accurate diagnosis of diabetes that contains a few additional factors that are involved for diabetes in addition to standard indicators such as blood glucose, body mass index (BMI), age, insulin, and so on. Garca-Ordás et al. (17) introduced an algorithm based on deep learning approaches to detect diabetes patients. Variational autoencoders (VAEs) can be used to add data and features, and a CNN can be used to classify the data. Alam et al. (18) diagnosed diabetes utilizing key variables and defined their relationships. Various techniques are utilized for diabetes clustering, prognosis, and association rules. The PIMA dataset was used by Naz and Ahuja (19) to diagnose diabetes. A neural network, Multilayer Perceptron, Logistic Regression, and Deep Learning are all effective classifications that attain 90–98% efficiency. Yuvaraj and SriPreethaa (20) presented machine learning techniques in hadoop clusters for diabetes diagnosis. The results reveal that machine learning techniques can correctly predict hyperglycemia. In their study, Hasan et al. (21) built a comprehensive system for diabetes prognosis that included components such as outlier exclusion, data normalization, extraction of features, K-fold cross-validation, and several machine learning (ML) models (k-nearest neighbor, Decision Trees, Randomized Forest, Xgboost, Bayesian Network, and Gradient boosting) and Lstm. An extensive study of the uses of deep learning in diabetes was published by Zhu et al. (22). Through the exploration, a lot of original scientific papers were found. Prediction models for impaired glucose tolerance in early pregnancy were formed by Liu et al. (23). They used machine learning to make these models. The training dataset was used to build a model for predicting risk based on information gathered at registration. The deep learning classification technique makes use of the ResNet v2 CNN architecture (24), which was trained on tiny patches taken from the entire ear endoscopies before being applied to the complete ear images. A total of four deep learning models were trained for autonomous ascribable diabetic retinopathy detection, dependent on whether or not two criteria were included: DR-related lesions and diabetic retinopathy staging (25). **Table 1** presents a summary of selected works.

Existing Research Gaps

Several deep learning methods based on word embedding have previously been successful in overcoming these limitations. Using deep learning algorithms, it is important to overcome the challenge of remembering extra information in order to make highly accurate diabetes predictions.

PROPOSED METHODOLOGY

It is vital to apply deep learning technology to incorporate current data and experience into a DSS in order to deal with this difficult decision-making challenge. Data, expertise, and models are incorporated into our DSS (see **Figure 1**) so that diabetic professionals may make diagnostic choices based on this information. To further involve the individuals, we sought the advice of medical specialists throughout the design process, as advised by (1). Considering the DSS's nature and the decision issue's complexity, we built the DSS according to Turban et al. (1) recommendations. The DSS has four main subsystems: data management, model management, knowledge-based management, and user interface.

Data Management System

A DSS uses databases and/or datasets to provide relevant data to the decision support system. DSS data can be received from local, public, and customized sources (1), as well as institutional sources. Our decision support system's data management module collects and stores data. In this investigation, the Pima Indians' Diabetes (PID) Data Set (26) was utilized as the data source. UCI's machine learning archive does have this dataset, and it is part of a larger set of data kept by the National Institutes of Health (27). This database contains information on women of Pima Indian

TABLE 1 | Summary of selected works.

References	Technique(s)	Results	Limitations
Butt et al. (9)	Randomized forest, multilayer perceptron, and regression models (LR)	MLP outperforms similar learners with an accuracy of 86%	Dimensionality reduction techniques not applied
Gupta et al. (10)	Support vector machine	Support vector machine did better than the naive Bayes model.	Lack of current comparability and consistency
Qawqzeh et al. (12)	Regression model	92% accuracy	Lack of comparison with the current methods
Pethunachiyar (13)	SVM classifier	Outperformed Naive Bayes, decision trees, and neural nets	There is no discussion of parameter estimation
Choubey et al. (14)	SVM, KNN, and NB	91% accuracy	Performance overhead due to incorporation of extensive feature engineering
Zhou et al. (15)	DTP model for glycemic control diagnosis	Promising results	Execution time needs to be further reduced
Garca-Ordás et al. (17)	Deep learning approach (CNN)	92.31% accuracy	Ensemble learning technique required for more better results
Alam et al. (18)	Association rules	92% accuracy	More effective preprocessing
Naz and Ahuja (19)	Multilayer Perceptron, Logistic Regression, and Deep Learning	90% accuracy	A pipeline of classifiers can produce efficient results
Yuvaraj and SriPreethaa (20)	Machine learning techniques in hadoop clusters	Outperformed baseline methods	Poor selection of predictors
Hasan et al. (21)	k-nearest neighbor, Decision Trees, Randomized Forest, Xgboost, Bayesian Network, Gradient boosting, and Lstm	LSTM exhibited better results (92% accuracy)	Parameter selection is affected by lack of efficient proper preprocessing techniques

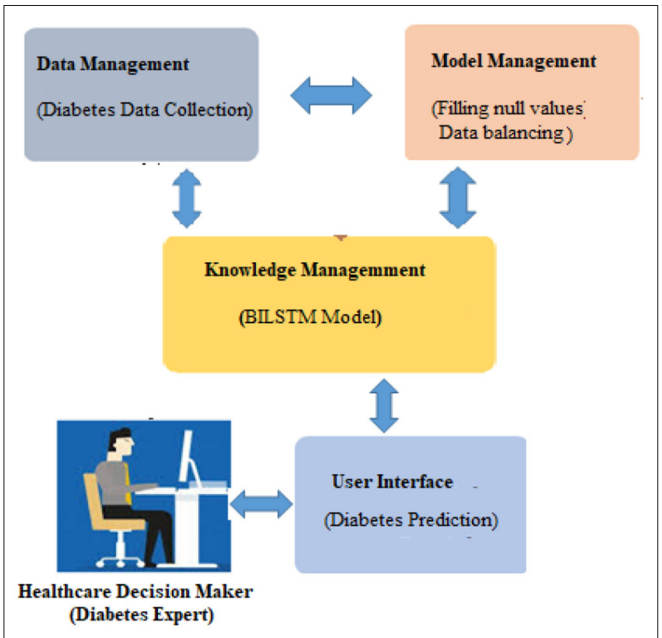


FIGURE 1 | Overview of the proposed system for predicting diabetes disease.

ancestry who were over the age of 20 at the time of data collection and who resided in the United States. The output data parameter accepts either a value of 0 or 1, with value 1 indicating a positive diabetic test and "0" indicating a negative diabetic diagnostic. An overview of the eight diagnostic attributes and their descriptions is provided in **Figure 2**.

How to Use Data

The spreadsheets are converted to CSV files. The "pd.read" command line option reads "csv". This is a key Panda tool. We separated the training and testing sets using the sklearn training (80%) and testing (20%) partition tool (2).

Train Set

During training, around 80% of the training set's data was used (3). The training set includes both result identifiers (dependent variables) and input factors (predictor variables).

Validation Set

Using validation data in the system, efficiency concerns such as overfitting and under fitting may be addressed. Thus, a 10% validation subset is employed for model assessment (2). Both manual and automated parameter changes are possible when using Keras. With the help of automated validation in this research, a more unbiased assessment of the proposed method may be made.

Test Set

In order to evaluate the algorithm's efficacy, the test set includes examples that have never been seen before. This method is applied after extensive usage of the training and testing sets. The model may be evaluated using the testing dataset (11). Ten percent of the test dataset was used, which had nothing to do with the training cases. The training data set is used when the model has been fully trained. It is then checked against real data for correctness. The data is divided into 90/10 ratios by the Scikit-train-test learning division, with 10% of the data being validation data. A validation set was used to

Attribute-id	attribute	description
Diabetes-f1	Age (years)	Age in years
Diabetes-f2	Body mass index	Body mass index (weight in kg/(height in m)2)
Diabetes-f3	Glucose	It takes two hours of oral blood glucose tolerance tests to get the plasma glucose concentration.
Diabetes-f4	Pregnancy	The number of pregnancies
Diabetes-f5	Blood pressure	Blood pressure measured in millimeters of mercury (diastolic, mm Hg).
Diabetes-f6	Skin Thickness	The skinfold thickness of the triceps (mm)
Diabetes-f7	insulin	Two-hour serum insulin (μ U/ml)
Diabetes-f8	Pedigree function of diabetes	Attribute that is used to predict diabetes.

FIGURE 2 | Parameters used to predict the likelihood of diabetes.

make adjustments to the model's parameters and then analyze the results.

Treatment of Data

The model is validated *via* 10-fold cross-validation. At each stage, we collect and keep ten replicas of the training instance. One last "holdout" model was examined in this case. We chose the version with the highest F1 score for the holdout sample.

Model Management System

It works as a data management system. It incorporates a modelbase of statistical and other algorithms that provide sophisticated analytics to DSS. An MMS applies models to DMS data to turn it into information. In order to create a reliable prediction model, the obtained health data must be properly pretreated. The data management system handles unbalanced datasets and null value substitution.

Unbalanced Data Set Management

The underlying dataset is significantly imbalanced and treats both groups unevenly. Two hundred sixty-eight cases (34.9%) are present in class one for a positive test, and five hundred sixty-one cases (65.1%) are present in class zero for a negative test. Whereas, when a model learns from skewed and unequal classed data, the result usually benefits the main class whereas the minor categories are neglected in the classification stage. This is viewed as a class imbalance issue (28). To solve this issue, we use a data processing sampling method called oversampling to equalize all class instances. Oversampling is a method of expanding small classes. Random upsampling merely duplicates minor examples to increase the unbalance percentage (28). This small group replication addition considerably enhanced the classification results. In both the T1:415 and T2:414 classes, the balanced dataset is treated in the same way once random oversampling is used. Instances in total: 829.

Handling Missing Values

In order to increase the model's prediction performance, missing variables should be filled in with accurate data (28). In order to fill in the blanks, one can use an average, choose a random item, or go back and use the prior tier's value as a reference. After settling on the third choice, we went ahead and updated all of the missing numbers with the appropriate higher value.

The Knowledge Management System

When used in conjunction with other DSS modules, this module may be able to provide the most up-to-date information to aid in the resolution of the issue at hand. After information has been found, acquired, and arranged, it must be transformed into knowledge. Data must be categorized, evaluated, and synthesized. The three key components of the proposed technique are: embedding layer-based data representation; bi-LSTM-based forward and backward context information saving; and softmax layer-based classification. With this numeric representation, the second module may encode features. Encoding the context of the data within a sequence using Bi-LSTM In the final module, classification is done using softmax activation (see **Figure 3**). Each component is described below.

Embedding Layer

A numeric array created by embedding items (categories) (vectors). Scalar word embeddings from discrete traits Using neural encoding to reduce attribute values and categorize has several benefits (29). Keras can embed. The diabetes dataset was encoded using attribute-level encoding matrices. Thirty-two Keras layers generate the data embedding vector. A two-dimensional embedding matrix had input data length and a word embedding dimension (feature matrix). This matrix was made by the embedding. The embedding matrices were then shifted.

The BILSTM Layer

The proposed method employs a deep neural network, specifically bidirectional long-short time memory (BiLSTM), to predict diabetes sickness, such as D1 (diabetes disease = yes) or D2 (diabetes disease = no). Long-term dependencies are learned using the Bi-LSTM layer. It assists in preserving the two prior and following contexts in encoded information. Instead of saving information from prior contexts, a single unidirectional LSTM only keeps data that was previously saved (30). Thus, Bi-LSTM is able to analyze encoded reviews in much more detail. Bi-LSTM uses forward and backward

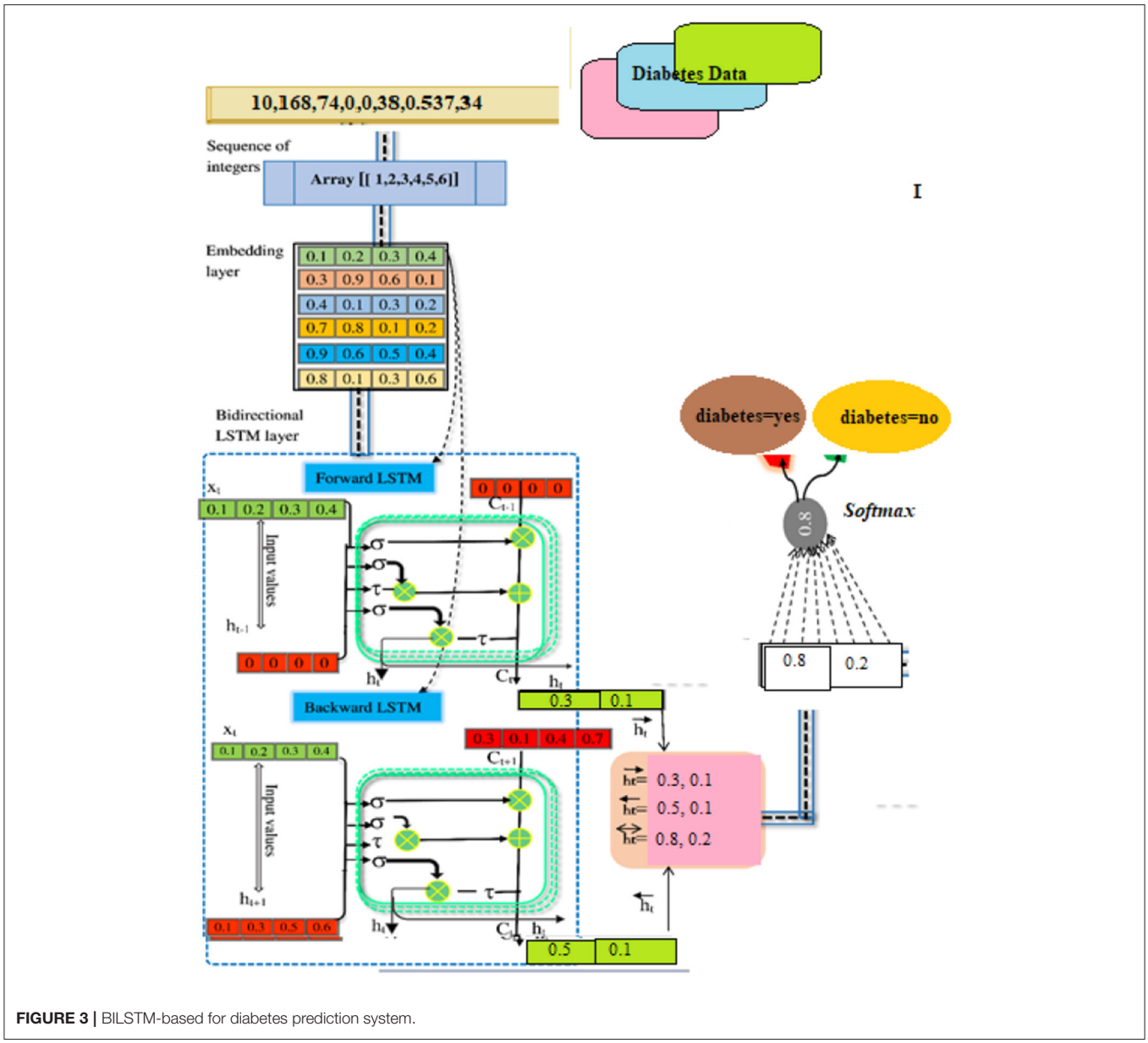


FIGURE 3 | BILSTM-based for diabetes prediction system.

LSTM to learn data's past and future context (2). The following formulas are used to determine the forward and reverse LSTM:

Formulae for forward LSTM

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (1)$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \quad (2)$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \quad (3)$$

$$c \sim t = \tau\left(W_c x_t + U_c h_{t-1} + b_c\right) \quad (4)$$

$$c_t = f_t \bigodot c_{t-1} + i_t \bigodot c \sim_t \quad (5)$$

$$h_t = o_t \bigodot \tau\left(c_t\right) \quad (6)$$

Formulae for backward LSTM

$$f_t = \sigma\left(W_f x_t + U_f h_{t+1} + b_f\right) \quad (7)$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t+1} + b_i\right) \quad (8)$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t+1} + b_o\right) \quad (9)$$

$$c \sim t = \tau\left(W_c x_t + U_c h_{t+1} + b_c\right) \quad (10)$$

$$c_t = f_t \bigodot c_{t+1} + i_t \bigodot c \sim_t \quad (11)$$

$$h_t = o_t \bigodot \tau\left(c_t\right) \quad (12)$$

SoftMax-Based Prediction

Afterward, SoftMax is being used to figure out how likely it is that target labels will be forecasted (i.e., the diabetes disease).

The formula (Equation 18) explains how to determine the net input value

$$d_i = \sum w_i l_i + b \tag{13}$$

"w" is the weight vector, whereas "l" stands for the input vector. "b" stands for "bias." We can find the SoftMax by plugging it into Equation (19).

$$softmax\ (d_i) = \frac{expd^{d_i}}{\sum_{n=1}^{m} \exp^{d_n}} \tag{14}$$

Applied Example

We performed a number of computations to predict diabetes based on the existing disease data. The BILSTM model's every stage is discussed in detail.

Data Preparation

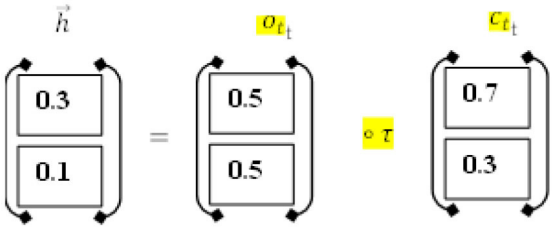
Our model predicts diabetes D1: diabetes Yes, or "D2: diabetes No" for a given patient instance in the illness dataset. Firstly, the illness data for the DL model is acquired through the instance selection module (31). The data was transformed into a matrix of indexes by a parser named Keras and transmitted to the embedding layer of the composite DL model for evaluation. The embedding layer converts each disease indicator into a vector containing streamed numbers. An example of a scalar embedding is [0.2 0.1 0.5 0.4], which encapsulates data about illness with the index [1]. In the end, the matrix packing looked somewhat like this: [0.36, 0.43, 0.85, 0.12], [0.52, 0.61, 0.11,0.25], [0.71, 0.22, 0.54, 0.47], [0.34, 0.48, 0.61, 0.39].

Extraction of Contextual Information

As input for this layer, a rectified feature map derived by the preceding neural network layer serves as input. In BILSTM layer calculations, the primary components are new candidate value ($c \sim_t$), output gate (o_t), forget gate (f_t), and input gate (i_t).

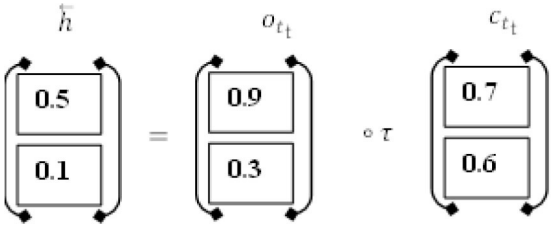
Hidden Layer No. 1

It includes the LSTM's current input (i_t) and prior state (h_{t-1}) The computations are done using Equations (1)–(6). Finally, the first hidden layer computes the "$\overrightarrow{h}$" hidden state (forward pass LSTM).



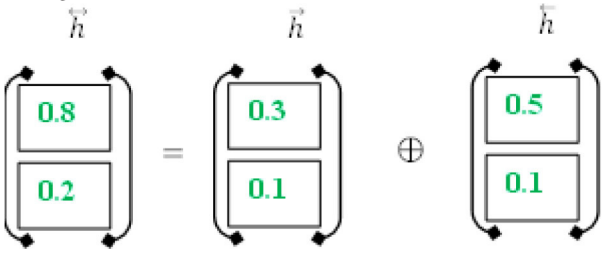
Hidden Layer No. 2

The present input (i_t) and the future state (h_{t+1}). make up the backward pass LSTM. The calculations are carried out by the use of the formulas (7) to (12). Finally, the succeeding layer is used to produce the hidden state $\overleftarrow{h}$ (backward pass LSTM).



BISLTM Outcome

In order to get the final BISLTM "$\overleftrightarrow{h}$" we combine the "$\overrightarrow{h}$" from the LSTM forward pass and the "$\overleftarrow{h}$" from the LSTM backward roll using Formula (5).



Diabetes Prediction

Using the SoftMax approach, it is determined how likely each of the labels "D1" and "D2" actually is. Formula (13) was used to determine the total input, as shown below:

For diabetes-yes, the class label for decision attribute 1 is "D1."

$$d_1 = l_1 \times w_2 + l_2 \times w_2 + b$$
$$d1 = 0.7 \times 0.8 + 0.5 \times 0.2 + 0.5$$
$$d_1 = 0.56 + 0.4 + 0.5 = 1.46$$

For diabetes-No, the class label for decision attribute 2 is "D2."

$$d_2 = l_1 \times w_2 + l_2 \times w_2 + b$$
$$d2 = 0.4 \times 0.8 + 0.4 \times 0.2 + 0.5$$
$$d2 = 0.32 + 0.08 + 0.3 = 0.7$$

To figure out the likelihood of each target class (D1, D2), the SoftMax function (14) is being used.

$$softmax\ (d_1) = \frac{\exp^{d_1}}{\sum_{i=1}^{n} \exp(d_1)}$$
$$softmax\ (d_1) = \frac{\exp^{1.46}}{\exp^{1.46} + \exp^{0.7}}$$
$$softmax\ (d_1) = \frac{4.305}{4.305 + 2.013} = \frac{4.305}{6.318} = 0.681$$

In the same way, the SoftMax function for the second class of diabetes predictions was made as well.

$$softmax\ (d_2) = \frac{\exp^{d_2}}{\sum_{i=1}^{n} \exp(d_2)}$$
$$softmax\ (d_2) = \frac{\exp^{0.7}}{\exp^{1.46} + \exp^{0.7}}$$
$$softmax\ (d_2) = \frac{2.013}{4.305 + 2.013} = \frac{2.013}{6.318} = 0.319$$

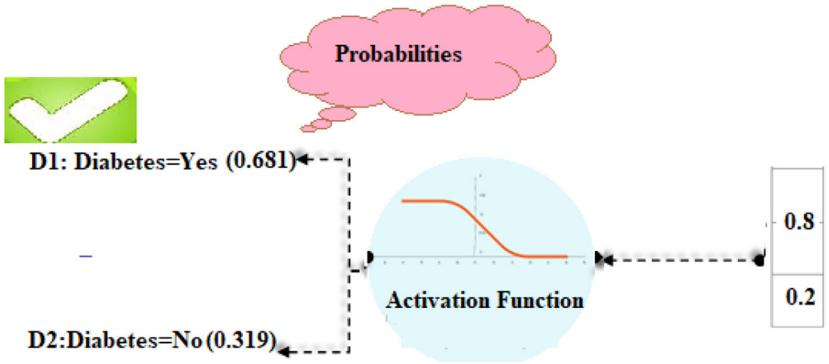


FIGURE 4 | Diabetes disease classification using the softmax function.

Algorithm 1 | Methodology of the proposed diabetes disease prediction model.

A. Data Input: As a csv file, import the diabetes disease labeled dataset.

B. Data Splitting: Spilt the Dataset into the Training and Testing Sets

D. Vocabulary Building: Build a vocabulary for mapping disease data to integers

E. Disease data stream transformation: Integerize each data stream

F. Preprocessing: Carry out data preprocessing

G. Data Balancing: Apply random oversampling to balance the dataset

H. Model Building and Classification: Create models using classification (BILSTM)

I. Performance evaluation: Prepare necessary measurements for each model to test predictive power

H. Make comparisons and choose:Comparing the results and finding the best model to use.

In our computations, class D1 diabetes was shown to have the highest likelihood (0.681). Based on such patient details, we can predict that the diabetic chance is "D1" (**Figure 4**).

Using pseudocode, **Algorithm 1** illustrates how the proposed method for forecasting diabetes illness works.

User Interface

The Keras package (1) provides a Python-based user interface for diabetes prediction. Diabetes forecasting software can help doctors and other healthcare providers predict diabetes. The information has been split into segments for simplicity of use and clarity of presentation. The software's body has three basic parts: data collection and preparation, classification and algorithm development, and diabetes diagnosis.

(i) ***Data Collection and Preprocessing Component:*** This component necessitates patient information. After that, the data is preprocessed on the backend. For every patient in the database, a new case patient identity is produced with the new and better data. Using preprocessing findings, a classifier and a model to predict the result of diabetic illness are constructed. The diabetes prediction component makes predictions about a patient's ailment based on the information they've supplied. A unique patient identifier is produced as soon as the required data is entered into the system. Every patient's precise ailment is therefore identified and tracked using this identification. **Figure 5** shows the screen that was used to collect and process the data for a patient registration input. At the software's server, the data is preprocessed in accordance with prior assumptions. As a result of our work, we have a layout that can be adapted to the individual requirements of each patient.

(ii) ***Classifier and Model Development:*** As demonstrated in **Figure 6**, a diabetic dataset may be utilized to train classifiers and build models. When a patient selects the "Model Training" tab, the screen below appears. It displays the imported data for training purposes. Clicking "Model Training" creates and trains a deep learning approach.

(iii) ***Predicting Diabetes:*** It is as simple as entering the necessary patient information and clicking the "Predict Diabetes" option on the site to predict diabetes. Whenever the "update training set" tab is pressed, a new training set is generated. After inputting the patient's disease information and pressing the "predict diabetes" button, the patient's findings are shown as "D1: diabetes Yes," "D2: diabetes No," and a predicted level of acceptance for each choice selected. As shown in **Figure 7**, the likely possibility of a diabetic illness diagnosis for a given set of criteria is "D1: diabetes Yes."

EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we will go through the findings from a series of experiments designed to address the questions posed in Section Introduction of this article.

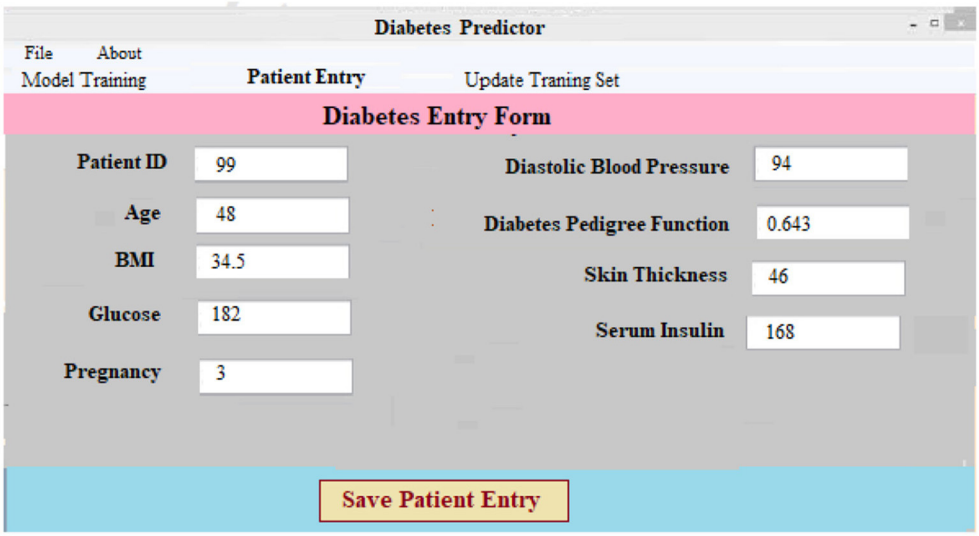
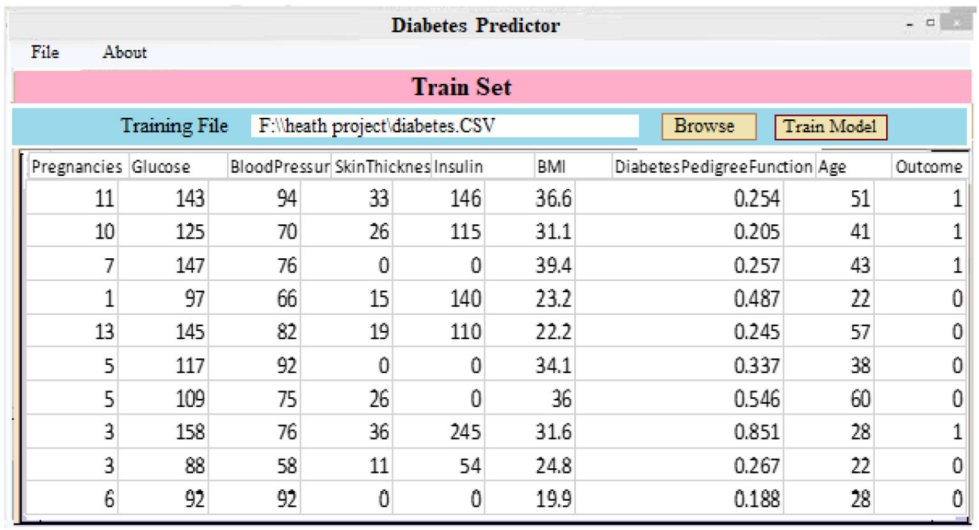


FIGURE 5 | Data entry form for diabetes prediction.



Pregnancies	Glucose	BloodPressur	SkinThicknes	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
11	143	94	33	146	36.6	0.254	51	1
10	125	70	26	115	31.1	0.205	41	1
7	147	76	0	0	39.4	0.257	43	1
1	97	66	15	140	23.2	0.487	22	0
13	145	82	19	110	22.2	0.245	57	0
5	117	92	0	0	34.1	0.337	38	0
5	109	75	26	0	36	0.546	60	0
3	158	76	36	245	31.6	0.851	28	1
3	88	58	11	54	24.8	0.267	22	0
6	92	92	0	0	19.9	0.188	28	0

FIGURE 6 | Create a model screen by loading training data.

Addressing Research Objectives

RO1: To Use the BiLSTM Deep Learning Model to Make Predictions About Diabetes Based on Patient Illness Data

We achieved study objective #1 by using multiple BILSTM algorithms for diabetes prediction by modifying the parameters of the recommended BILSTM model. Additionally, there were several other epochs and filtering methods in use. The algorithm has three hidden layers as well as a number of batch sizes and epochs. The number of vocabulary vectors was 62, and the embedding dimension was 128 with the SoftMax activation function being employed (8, 16, and 32). **Table 2** shows the accuracy, recall, and F-score of several BILSTM models. It is 93% accurate with the following parameters: filter number 8, filer size 280, unit size 2, "f1 score" of 92%, recall of 92%, and precision of 92%.

Data balancing greatly increases efficiency, as seen in **Table 3**, compared to not employing data balancing. Using the experimental data, the suggested model may be used to accurately forecast diabetes in real-world contexts.

Table 4 shows the results of 10 BILSTM trials with varied parameter values. We compared the accuracy of each model. The BILSTM-10 model had the greatest accuracy, with eight filter sizes, 16 filters, and 10 LSTM units.

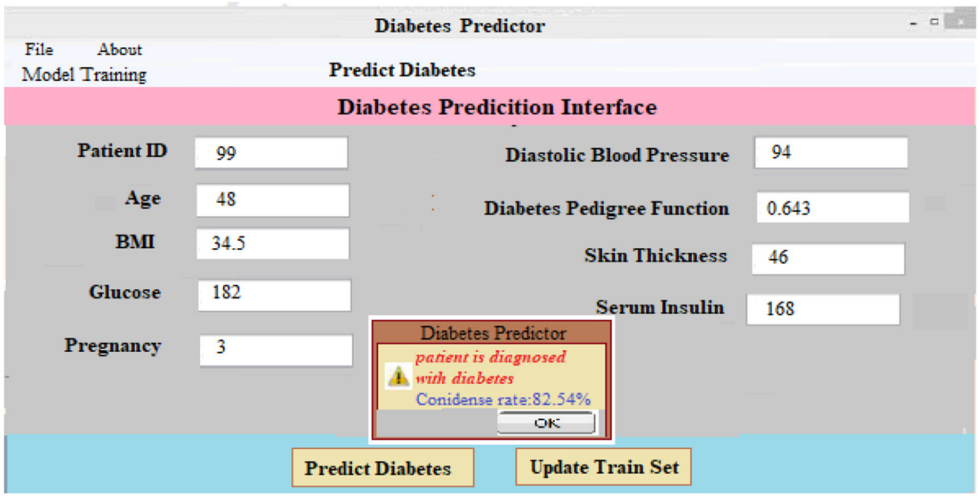


FIGURE 7 | Diabetes prediction interface.

Computational Complexity

Since the input data is routed *via* two LSTM layers, the computational cost of a conventional LSTM model per each stage with a gradient descent optimizer is O (W), where W is the maximum number of variables (21). However, despite the fact that BiLSTM has high computational complexity, it is successful in reducing the volume and complexity of the feature space. BiLSTM takes advantage of the data's inherent properties. Using drop out, we can cut down on the number of features and make sure the model doesn't fit even more than it should.

RO2: Comparison of the BiLSTM Model for Diabetes Prediction With Classical Machine Learning and Deep Learning

The BILSTM results for diabetic disease prediction were contrasted with those from several traditional machine learning approaches as well as deep learning techniques in order to address the second study objective.

Machine Learning vs. Proposed System (BILSTM)

In order to evaluate the suggested approach (BILSTM) with other common machine learning methods, data from patients was employed. Feature representation techniques such as TF-IDF and CountVectorizer are used in machine learning. **Table 5** shows the results of the performance evaluations. The results are summarized below.

- **BILSTM vs. KNN**

  To see how well the suggested BILSTM model stacks up against other machine learning techniques, we used the K-nearest neighbors approach. **Table 5** shows the results of the analysis.
- **BILSTM vs. DT**

  The purpose of this experiment was to compare a BILSTM model against a traditional machine learning classifier (DT). **Table 5** shows that the precision (0.81), recall (0.78), F1-score (0.80), and accuracy (0.78) of DT classifiers were all poorer (80%).
- **BILSTM vs. SVM**

  The results from the BILSTM model were found to be more effective than those from the SVM classification algorithm. **Table 5** displays a lower F1-score (0.79), a lower recall (0.81), and a lower precision (0.79).
- **BILSTM vs. NB**

  BILSTM was tested against a Nave Bayes (NB) classifier in this experiment. **Table 5** shows a lower F1-score (0.70%), lower recall (0.70%), and lower precision (0.70%).

TABLE 2 | BILSTM deep learning models' accuracy, recall, and f1-score.

Model	Precision (%)	Recall (%)	F-score (%)
BILSTM-1	0.74	0.75	0.74
BILSTM-2	0.83	0.77	0.75
BILSTM-3	0.85	0.82	0.80
BILSTM-4	0.86	0.84	0.83
BILSTM-5	0.89	0.85	0.84
BILSTM-6	0.88	0.88	0.87
BILSTM-7	0.89	0.89	0.88
BILSTM-8	0.90	0.89	0.89
BILSTM-9	0.92	0.90	0.90
BILSTM-10	0.93	0.92	0.92

TABLE 3 | Performance of the BILSTM models with and without balancing data.

Performance measure	Without data balancing	With data balancing
Accuracy (%)	82	93.07
Precision (%)	82	93
Recall (%)	82	92
F1-score	88	92

TABLE 4 | The BILSTM models' accuracy, test loss, and training time.

Model	Test accuracy (%)	Test loss	Train time (s)
BILSTM1	81.23	0.78	18
BILSTM2	82.21	0.86	5
BILSTM3	83.71	1.04	17
BILSTM4	84.67	1.13	15
BILSTM5	85.98	0.92	6
BILSTM6	87.47	0.91	13
BILSTM7	88.21	1.11	11
BILSTM8	88.36	0.80	13
BILSTM9	89.23	0.91	15
BILSTM10	92.15	0.82	10

TABLE 5 | Machine learning classifiers vs. proposed model (BILSTM).

ML modal	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
KNN	78	79	80	79
DT	81	80	80	80
SVM	82	79	81	79
NB	72	70	70	70
Proposed (BILSTM)	93.07	93	92	92

TABLE 6 | BILSTM vs. other DL models.

DL model	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
LSTM	83.36	84	83	83
CNN	83.22	81	80	81
RNN	81.11	80	81	80
Proposed (BILSTM)	93.07	93	92	92

Deep Learning vs. Proposed Technique (BILSTM)

For the purpose of accurately predicting diabetes based on patient data, the suggested method is compared to other deep learning techniques, such as long/short-term memory (LSTM), convolutional neural networks (CNN), and recurrent neural networks (RNN). **Table 6** summarizes the results.

- **LSTM vs. Proposed BILSTM**

  We compared the BILSTM model's performance to that of a single LSTM model throughout this research. The LSTM model has the lowest precision, recall, F1-score, and accuracy among the models shown in **Table 6**.
- **CNN vs. Proposed BILSTM**

  We wanted to see if the suggested BiLSTM method outperformed the CNN model in this trial. **Table 6** shows that the CNN model performed poorly in terms of accuracy, precision, recall, F1 score, and precision.
- **RNN vs. Proposed BILSTM**

  We conducted this test in order to ascertain which method was the most effective. In **Table 6**, the RNN model's precision, recall, F1-score, and accuracy were found to be worse than expected.

TABLE 7 | Comparison of the BILSTM model with other studies.

References	Technique	Results
Butt et al. (9)	Predicting diabetic illness with machine learning	Acc: 85%
Gupta et al. (10)	Predicting diabetic illness with machine learning	Acc. 88%
Proposed (BILSTM)	Predicting diabetes with deep learning (BILSTM)	93.07

RO3: Comparison of Proposed Method's Effectiveness to Baseline Research for Predicting Diabetic Patients

For the third study question, we compared the proposed BILSTM model's efficacy to similar studies. The suggested system is compared to numerous benchmarking approaches to assess its efficiency. It compares our suggested BILSTM approach to a baseline study and shows that it outperforms the latter (**Table 7**). An exhaustive review of published approaches is challenging for numerous reasons. With so many distinct datasets, it was difficult to compare these methods.

First Study

Butt et al. (9) proposed a supervised ML model for diabetes prediction based on patient data. Data from diabetic patients was collected using a variety of machine learning techniques. The experimental findings demonstrate that the model's performance is unsatisfactory (accuracy: 88%).

Second Study

A ML-based approach for diabetes prediction has been suggested. They used a range of machine learning approaches to analyze diabetes data sets. Combining a DL model with a more effective strategy for data balancing may improve the model's performance.

Work Proposed (Our Model)

A deep neural network is employed in the proposed DL-based diabetes prediction method. The experimental results (**Table 7**) show that the predictor attributes (**Table 7**) chosen have a significant impact on the predicted (target) variable. The main reason for our success in predicting diabetic diseases is the integration of data balancing and the BILSTM deep learning model. With the help of the BILSTM layer, context data may be preserved.

Analyzing Results

Experts' predictions are compared to the forecast provided by the proposed technique, and the proposed method's performance is evaluated. The first 12 patients' workflows are shown in **Table 8**.

TABLE 8 | The human expert's prognosis vs. the suggested system's.

Suspected diabetic patient	Diagnosis by diabetes expert	Prediction by BILSTM model (proposed)
1	Diabetes = yes	Diabetes = yes
2	Diabetes = yes	Diabetes = yes
3	Diabetes = no	Diabetes = no
4	Diabetes = yes	Diabetes = no
5	Diabetes = yes	Diabetes = yes
6	Diabetes = yes	Diabetes = yes
7	Diabetes = yes	Diabetes = yes
8	Diabetes = no	Diabetes = no
9	Diabetes = yes	Diabetes = yes
10	Diabetes = no	Diabetes = no
11	Diabetes = yes	Diabetes = yes
12	Diabetes = yes	Diabetes = yes

Threats to External Validity

As indicated in Section Addressing Research Objectives, the suggested strategy was evaluated internally to assure model stability, and two extra datasets were acquired to externally support the proposed methodology. We collected two more datasets after conducting internal validation of the suggested technique to guarantee the validity of the strategy.

Dataset 2

This dataset comes from the University of Virginia School of Medicine's Department of Medicine (32). It has 1,046 occurrences in two classes. Based on comparisons with the Pima Indians Diabetes Dataset and clinical experience, we chose 12 key features from 19 initial attributes that we thought were most important to look for.

Dataset 3

This data set was compiled *via* responses to an internet questionnaire. There are 14 variables in the questionnaire that we came up with: age, gender, being pregnant, family history, BMI, sleeping habits, quality of sleep, snoring and other snoring-related behaviors, appetite cues, tobacco, drinks, hypertension, sugar levels, and blood sugar. Sixty-eight positive and 316 negative cases make up the dataset, which is broken down into two groups. This data set of people who live in Pakistani let us test our model in a real-world setting.

By comparing it to the multiple datasets provided in this section, we demonstrate that the proposed model for diagnosing diabetes is both effective and exact. In order to evaluate the suggested approach, classifiers are created and tested on the given data sets. On the other hand, models that have been trained on the primary dataset (dataset 1) will be tested on the two new datasets that have been set up.

Table 9 provides a summary of the findings. Both the KNN and LSTM baseline techniques were tested against the proposed model (BILSTM). An average of 82% of the compared techniques (KNN and LSTM) were outperformed by the proposed solution. This study's findings validate the proposed model and its potential to enhance classification accuracy.

TABLE 9 | The external validation of the proposed method.

Dataset	Precision	Recall	Accuracy
Dataset 2	0.723 (KNN)	0.718 (KNN)	0.751(KNN)
	0.761 (LSTM)	0.752 (LSTM)	0.766 (LSTM)
	0.855 (Proposed)	0.819 (Proposed)	0.811 (Proposed)
Dataset 3	0.718 (KNN)	0.817 (KNN)	0.781 (KNN)
	0.726 (LSTM)	0.761 (LSTM)	0.761 (LSTM)
	0.813 (Proposed)	0.834 (Proposed)	0.826 (Proposed)

CONCLUSIONS AND FUTURE WORK

Because of the massive increase in healthcare content, the collection and analysis of such data to identify diabetes disease in patients has become increasingly significant. In order to do this, an effective DL-based DSS model was developed and applied. Get benchmarks for data collection, preprocess, and then use a deep neural network (BILSTM) to forecast diabetes are three components of this model. In addition, the balanced and unbalanced data sets were used in subsequent tests. A BILSTM model was also used to estimate the likelihood of diabetes developing in the future. When compared to previous attempts, the findings are positive. The proposed method does have a few apparent limitations, such as the use of embedding rather than a pre-trained model. In the future, the use of diabetes data sets from many domains (e.g., patient data from several domains) with pre-trained algorithms like word2vec or Fasttext may be investigated.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

DA: conceptualization, supervision, project administration, and funding acquisition. DA and OR: methodology and formal analysis. FS and JA: software. DA and JA: validation. OR: investigation. MA: resources and writing—original draft preparation. FS and OR: data curation. JA: writing—review and editing. MA and JA: visualization. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant No. (611-008-D1434).

ACKNOWLEDGMENTS

The authors acknowledge with thanks to DSR technical and financial support.

REFERENCES

1. Turban E., Aronson, JE, Liang, TP, MacCarthy RV. *Decision Support Systems and Intelligent Systems*. Upper Saddle River, NJ: Pearson Prentice-Hall (2005).
2. Ahmad H, Asghar MU, Asghar MZ, Khan A, Mosavi AH. A hybrid deep learning technique for personality trait classification from text. *IEEE Access.* (2021) 9:146214–32. doi: 10.1109/ACCESS.2021.3121791
3. Khattak A, Habib A, Asghar MZ, Subhan F, Razzak I, Habib A. Applying deep neural networks for user intention identification. *Soft Comput*. (2021) 25:2191–220. doi: 10.1007/s00500-020-05290-z
4. Rose DC, Sutherland WJ, Parker C, Lobley M, Winter M, Morris C, et al. Decision support tools for agriculture: towards effective design and delivery. *Agric Syst.* (2016) 149:165–74. doi: 10.1016/j.agsy.2016.09.009
5. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* (2017) 22:1589–604. doi: 10.1109/JBHI.2017.2767063
6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* (2018) 19:1236–46. doi: 10.1093/bib/bbx044
7. The Checkup. *Diabetes Statistics: Read the Facts*. (2020). Available online at: https://www.singlecare.com/blog/news/diabetes-statistics/ (accessed December 20, 2021).
8. *Idf Diabetes Atlas*. 10th ed. Available online at: https://diabetesatlas.org/ (retrieved January 6, 2022).
9. Butt UM, Letchmunan S, Ali M, Hassan FH, Baqir A, Sherazi HHR. Machine learning based diabetes classification and prediction for healthcare applications. *J Healthcare Eng.* (2021) 2021:9930985. doi: 10.1155/2021/9930985
10. Gupta S, Verma HK, Bhardwaj D. Classification of diabetes using naïve bayes and support vector machine as a technique. In: Sachdeva A, Kumar P, Yadav OP, Garg RK, Gupta A, editors. *Operations Management and Systems Engineering*. Singapore: Springer (2021). p. 365–76. doi: 10.1007/978-981-15-6017-0_24
11. Alghazzawi D, Bamasaq O, Ullah H, Asghar MZ. Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection. *Appl Sci.* (2021) 11:11634. doi: 10.3390/app112411634
12. Qawqzeh YK, Bajahzar AS, Jemmali M, Otoom MM, Thaljaoui A. Classification of diabetes using photoplethysmogram (PPG) waveform analysis: logistic regression modeling. *Biomed Res Int.* (2020) 2020:3764653. doi: 10.1155/2020/3764653
13. Pethunachiyar GA. Classification of diabetes patients using kernel based support vector machines. In: *2020 International Conference on Computer Communication Informatics (ICCCI)*. Coimbatore: IEEE (2020). p. 1–4. doi: 10.1109/ICCCI48352.2020.9104185
14. Choubey DK, Kumar M, Shukla V, Tripathi S, Dhandhania VK. Comparative analysis of classification methods with PCA and LDA for diabetes. *Curr Diabetes Rev.* (2020) 16:833–50. doi: 10.2174/1573399816666200123124008
15. Zhou H, Myrzashova R, Zheng R. Diabetes prediction model based on an enhanced deep neural network. *EURASIP J Wireless Commun Network.* (2020) 2020:1–13. doi: 10.1186/s13638-020-01765-7
16. Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. *Proc Comput Sci.* (2019) 165:292–9. doi: 10.1016/j.procs.2020.01.047
17. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. *Comput Methods Prog Biomed.* (2021) 202:105968. doi: 10.1016/j.cmpb.2021.105968
18. Alam TM, Iqbal MA, Ali Y, Wahab A, Ijaz S, Baig TI, et al. A model for early prediction of diabetes. *Inform Med Unlocked*. (2019) 16:100204. doi: 10.1016/j.imu.2019.100204
19. Naz H, Ahuja S. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J Diabetes Metab Disord.* (2020) 19:391–403. doi: 10.1007/s40200-020-00520-5
20. Yuvaraj N, SriPreethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. *Cluster Comput.* (2019) 22:1–9. doi: 10.1007/s10586-017-1532-x
21. Hasan MK, Alam MA, Das D, Hossain E, Hasan M. Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access.* (2020) 8:76516–31. doi: 10.1109/ACCESS.2020.2989857
22. Zhu T, Li K, Herrero P, Georgiou P. Deep learning for diabetes: a systematic review. *IEEE J Biomed Health Inform*. (2020) 25:2744–57. doi: 10.1109/JBHI.2020.3040225
23. Liu H, Li J, Leng J, Wang H, Liu J, Li W, et al. Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China. *Diabetes Metab Res Rev.* (2021) 37:e3397. doi: 10.1002/dmrr.3397
24. Pfister M, Stegmann H, Schützenberger K, Schäfer BJ, Hohenadl C, Schmetterer L, et al. Deep learning differentiates between healthy and diabetic mouse ears from optical coherence tomography angiography images. *Ann N Y Acad Sci.* (2021) 1497:15. doi: 10.1111/nyas.14582
25. Wang Y, Yu M, Hu B, Jin X, Li Y, Zhang X, et al. Deep learning-based detection and stage grading for optimising diagnosis of diabetic retinopathy. *Diabetes Metab Res Rev.* (2021) 37:e3445. doi: 10.1002/dmrr.3445
26. Pima Indians Diabetes Database. *Data.World*. Available online at: https://data.world/data-society/pima-indians-diabetes-database (retrieved January 6, 2022).
27. UCI Machine Learning Repository. *Diabetes Data Set*. Available online at: https://archive.ics.uci.edu/ml/datasets/diabetes (retrieved January 6, 2022).
28. Khan AS, Ahmad H, Asghar MZ, Saddozai FK, Arif A, Khalid HA. Personality classification from online text using machine learning approach. *Int J Adv Comput Sci Appl.* (2020) 11:460–71. doi: 10.14569/IJACSA.2020.0110358
29. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmanna K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Human Comput.* (2021). doi: 10.1007/s12652-021-03154-4
30. Gadekallu TR, Gao XZ. An efficient attribute reduction and fuzzy logic classifier for heart disease and diabetes prediction. *Recent Adv Comput Sci Commun.* (2021) 14:158–65. doi: 10.2174/2213275911666181030124333
31. Bhattacharya S, Maddikunta PKR, Hakak S, Khan WZ, Bashir AK, Jolfaei A, et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimed Tools Appl.* (2020). doi: 10.1007/s11042-020-09988-y
32. Wu H, Yang S, Huang Z, He J, Wang X. Type 2 diabetes mellitus prediction model based on data mining. *Inform Med Unlocked.* (2018) 10:100–7. doi: 10.1016/j.imu.2017.12.006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Rabie, Alghazzawi, Asghar, Saddozai and Asghar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fpubh.2022.858327

Retinal Vessel Extraction *via* Assisted Multi-Channel Feature Map and U-Net

Surbhi Bhatia[1]*, *Shadab Alam*[2], *Mohammed Shuaib*[2], *Mohammed Hameed Alhameed*[2], *Fathe Jeribi*[2] *and Razan Ibrahim Alsuwailem*[1]

[1] Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Hofuf, Saudi Arabia, [2] College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia

Early detection of vessels from fundus images can effectively prevent the permanent retinal damages caused by retinopathies such as glaucoma, hyperextension, and diabetes. Concerning the red color of both retinal vessels and background and the vessel's morphological variations, the current vessel detection methodologies fail to segment thin vessels and discriminate them in the regions where permanent retinopathies mainly occur. This research aims to suggest a novel approach to take the benefit of both traditional template-matching methods with recent deep learning (DL) solutions. These two methods are combined in which the response of a Cauchy matched filter is used to replace the noisy red channel of the fundus images. Consequently, a U-shaped fully connected convolutional neural network (U-net) is employed to train end-to-end segmentation of pixels into vessel and background classes. Each preprocessed image is divided into several patches to provide enough training images and speed up the training per each instance. The DRIVE public database has been analyzed to test the proposed method, and metrics such as Accuracy, Precision, Sensitivity and Specificity have been measured for evaluation. The evaluation indicates that the average extraction accuracy of the proposed model is 0.9640 on the employed dataset.

Keywords: multichannel, retinal vessels, retinopathy, U-Net, Cauchy distribution

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Haider Khalaf Jabbar, Imam Jafar Al-Sadiq University, Iraq

***Correspondence:**
Surbhi Bhatia
sbhatia@kfu.edu.sa

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 January 2022*
Accepted: *04 February 2022*
Published: *17 March 2022*

Citation:
Bhatia S, Alam S, Shuaib M, Hameed Alhameed M, Jeribi F and Alsuwailem RI (2022) Retinal Vessel Extraction via Assisted Multi-Channel Feature Map and U-Net. Front. Public Health 10:858327. doi: 10.3389/fpubh.2022.858327

INTRODUCTION

Retinal vessel detection and identifying blood vessel properties, including diameter, shape, bifurcation, and tortuosity, mainly aim to diagnose various eye abnormalities and retinopathies (1). At the same time, the evaluation of bifurcation points and tortuosity assist in diagnosing cardiovascular ailments. Also, analyzing width can result in the prevention of retinopathies induced by hypertension (2). Besides, glaucoma prevention is only possible if fundus retinal images are analyzed regularly to diagnose any abnormal modification of blood vessels around the optic disk (3). Diabetic retinopathy is another disease imposed by blood vessel construction, distribution changes, and fresh vessel evolution. Late diagnosis may also result in adult blindness (4). These critical medical applications highlight the importance of vessel extrication from retinal fundus pictures for ophthalmologists in diagnosing diseases and acting properly for their patients.

Unfortunately, with such a significant role the retinal vessel detection, blood vessels are not visible easily in retinal images taken by the fundus. The red background of the retina alongside the red color of retinal vessels is combined with fading small vessels spread around the retina. Therefore, manual detection of blood vessels is considered a subjective task requiring specific expertise while time-consuming (5). The vessel's width varies subject to the breadth of the vessel and the image resolution. The retinal boundary, optic disc, and diseases such as cotton wool patches, bright and black lesions, and exudate can all be seen in ocular fundus images (3).

Image processing methods have improved the recognition of blood vessels by changing the intensity of images and normalizing the color distribution. These techniques can enhance the reading of retinal images; however, fully automatic extraction requires understanding the properties of the vessel and detecting them in the noisy background of retinal images at which inadequate contrast regions are distrusted non-uniformly. It has been more than three decades where Computer-Aided-Detection (CAD) methods have been employed to diagnose retinopathies. Unfortunately, most traditional methods are not well-suited for recent, accurate sensors. For instance, since the resolution of fundus scanners is increased, standard image-processing methods show their drawback in detecting small faded vessels. Briefly said, the traditional systems underperform the detection due to variations in image properties due to different capture devices that influence the intensity and resolution of retinal images; therefore, machine learning-based approaches are more suitable (6).

For vascular segmentation of retinal pictures, various machine learning (ML) approaches and procedures are available. These methods are divided into supervised and unsupervised approaches. Support vector machine (SVM), neural network, and classification based methods are examples of supervised methods. Matching filter, vessel tracking, and mathematical morphology methods are all examples of unsupervised approaches (7).

The remaining current paper is divided into three sections. In Section Literature Review, a brief review will be conducted on the approaches targeting the objective of the work *via* various techniques. The section is wrapped up by highlighting the importance of fusion approaches taking novel DL methods with successful traditional approaches at once. In the third section, the main idea of the research is proposed, while the user data set and the structure of the employed DL is detailed. Later in Section Results and Discussions, results will be reported, and evaluations will be carried out to identify the strengths and limitations of the suggested approaches compared with the other similar methods. The conclusion is then followed in Section Conclusion and Future Work to wrap up the discussed methodology.

LITERATURE REVIEW

A broad look at the Literature review covering retinal blood vessel detection methodologies reveals that the current methodologies can be categorized into three classes (8) template matching based, tracking-based, and classifying based. In template matching approaches, a kernel is selected, which can be fitted well to a vessel. This kernel is later convoluted with each image. At the final stage, pixels with the highest responses are collected *via* a proper thresholding method to provide a binary map of vessel/background. A Gaussian-shaped curve proposed by Chaudhuri et al. (9) is used to highlight the vessels based on the intensity distribution of the vessel's cross-section (as illustrated in **Figure 1**).

Later improvement has been proposed by Al-Rawi et al. (10) to tune the parameters to increase the Sensitivity and Accuracy and decrease false positive detection. Gaussian, Cauchy, and Gabor distributions have been employed with the same purpose (3).

In contrast to the previous two approaches in which a template of the vessels plays an essential role in the detection, tracking approaches use a model to trace vessels *via* seed growing. Each of the vessels can be traced separately from a starting point (which can be an ending point of the parent vessel) to the ending point at which the vessel is branched. Considering the contrast between vessels and other retina background intensity, Wu et al. (11) determined the mid-line of a vessel from a starting point inside it. Having calculated the vessel's width, tracing the orientation is taken into place automatically by assuming the vessel boundaries are parallel with the middle line. Several discussions have been conducted to detect the initial seed point *via* manual annotation or automatic template matching methods (7). For instance, the methodology proposed by Dizdaro et al. used level-set thresholding to identify an initial seed (12). The work carried out by Zhang et al., which applied active contours started from a seed found *via* Hessian vessel response (13), can be categorized in the second explained group using a tracking approach to detect vessels. Ali et al. used B-COSFIRE filter blending with adaptive thresholding for binarization of retinal blood vessels (14).

In classifying approaches, mostly ML techniques are taken into consideration. ML methods focus on two sets of features: common between pixels belonging to vessels and those that discriminate between background and vessel pixels. Provision of such pixels can increase the accuracy of a model being trained by artificial intelligence (AI) (15, 16).

On the one hand, practical ML approaches with supervised methods target pixel-based feature maps as training to identify vessel patterns. Various methodologies have been proposed to detect vessels using k-Nearest Neighbors (17), Decision Trees (18), SVM (19), and Neural Networks (20, 21). On the other hand, unsupervised methods employ rule-based algorithms, including filters, gradients, and thresholds for the same classification purpose. For instance, the work presented in (22) employs a neural network trained to model a matched filter.

Traditional ML algorithms and recent deep-learning methods can be listed as the third set of approaches. Recent studies on DL comprise two perspectives, in which the first one addresses retinal vessel detection *via* pixel-level classification. The later perspective uses DL to achieve semantic segmentation between the two classes of pixels. An instance can be the pipeline of methodologies proposed by Wang et al. (23). They extract a set

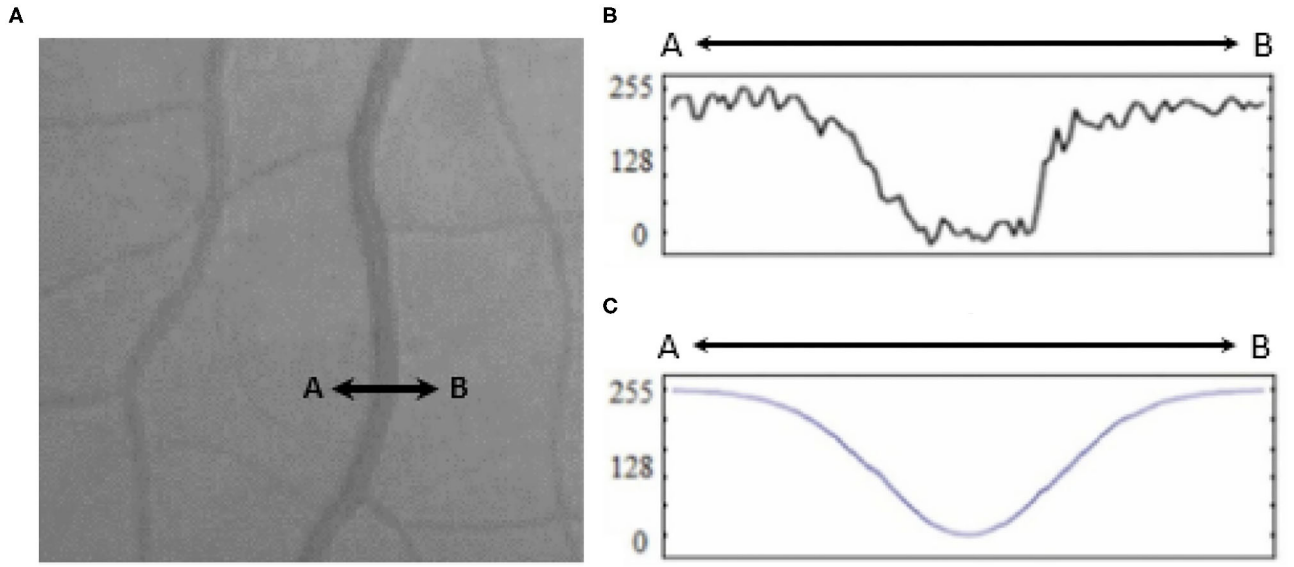


FIGURE 1 | (A) Retinal vessel cross-section (A to B), **(B)** intensities of the cross-section, and **(C)** fitted Gaussian curve proposed by Chaudhury et al. (9).

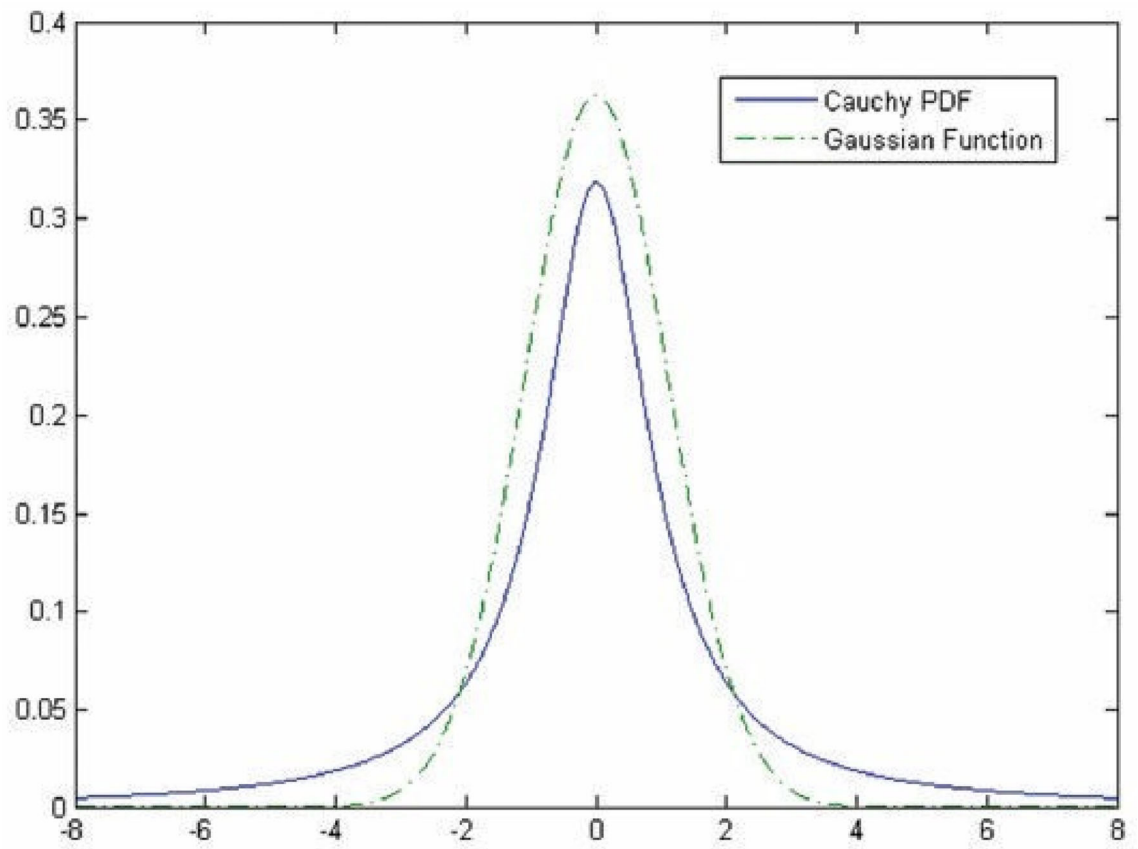


FIGURE 2 | Comparison of the Gaussian and the Cauchy curves for generating a template for retinal vessels [The image is adapted from (34)].

TABLE 1 | Related works.

References	Used techniques	Main idea of paper
(24)	Unsupervised hierarchical feature learning and ensemble	Suggested using an ensemble of stacked denoising autoencoders. Subsequently, the final decision was made *via* voting of a SoftMax layer over autoencoder outputs.
(29)	Deep CNN and nearest neighbor search	Suggested the employment of CNN to extract binary masks followed by a generalized particle filtering technique for vascular structure identification from medical images aiming at retinal vessel tree extraction *via* a probabilistic tracking structure.
(30)	Conditional Random Field (CRF), multi-level CNN	Carried out the vessel detection by Conditional Random Field (CRF) developed based on a multi-level CNN model.
(34)	Cauchy PDF and Gaussian function	Showed that the Cauchy probability distribution function (PDF) could be a model cross-section of vessels more accurate than the Gaussian.
(33)	Gaussian matched filter with U-Net	To use Gaussian matched filter in the preprocessing stage with U-Net CNN, aiming to improve the detection of vessels.
(35)	Y-net	Presented a Residual Y-net design for retinal vessel detection inspired by U-net that help in Diabetic detection.
(36)	U-Net	A review of U-Net and its variant techniques for image segmentation and specifically highlight their applications in retinal fundus image segmentation.
(37)	Multiscale matched filter, U-Net	Proposes a model that combines a multiscale matched filter with a U-Net that has been tested on various available public datasets.
(38)	Multiscale, texture feature-based, Segmentation-based, CNN, Ensemble Learning approaches	Describe different types of ML models for glaucoma detection from fundus images like multiscale, texture feature-based, Segmentation-based, CNN, Ensemble Learning approaches in detail.
(39)	Dense U-net	Proposed a structure using Dense U-net and the patch-based learning approach for clinical applications.

of required features using DL and latterly distinguish each pixel *via* Random Forests.

Ensemble learning has also been used in several investigations. Lahiri et al. (24) suggested using an ensemble of stacked denoising autoencoders. Subsequently, the final conclusion is made *via* voting of a SoftMax layer over autoencoder outputs. In another work, Maji et al. (25) suggested an ensemble of 12 deep Convolutional neural networks (CNN). The final decision is made *via* performing averaging of the output result of the decision over each pixel. Few studies have targeted DNN and formulated the objective into a pixel-to-pixel classification (26, 27); however, these approaches are underperforming the performance by picking each pixel and deciding the proper class in terms of training or testing. A practical solution has been proposed by Dasgupta et al. (28) in which both training and testing phases are carried out over small retinal patches with an end-to-end perspective. These patches are individual from each other, and hence they can be processed simultaneously.

Wu et al. (29) suggested employing a CNN to extract binary masks followed by a generalized particle filtering technique aiming at retinal vessel tree extraction *via* a probabilistic tracking framework. Concerning long-range collaborations between pixels, Fu et al. (30) carried out the vessel detection by Conditional Random Field (CRF) developed based on a multi-level CNN model.

Hybrid methodologies can be suggested *via* a combination of the approaches mentioned above. For instance, the template matching approach can be utilized in the third approach to provide more features for classifying vessels. Studies combine these approaches in sequential order, as template matching has been considered in the preprocessing stage to highlight features that will be used later by the ML approach to discriminate pixels belonging to either retinal tissue or vessel (31, 32). For instance, the work proposed by Gao et al. (33) combines Gaussian matched filter (as a preprocessing stage) with a U-Net CNN to improve vessel detection. In other studies, Zolfagharnasab et al. (34) showed that the Cauchy probability distribution function (PDF) could be a model cross-section of vessels more accurate than the Gaussian. The comparison of Gaussian and the Cauchy curves for generating a template for retinal vessels has been shown in **Figure 2**.

In this way, the conducted study aims to suggest two main contributions in the pipeline of the work proposed by Gao et al. First, supported by the investigation conducted in (33), we will employ Cauchy PDF instead of Gaussian. Second, a novel feature map will be prepared instead of raw pixel intensities by proposing a new three-channel image. Roy and Sharma presented a Residual Y-net design for retinal vessel detection inspired by U-net that help in Diabetic detection (35). Siddique et al. gave a review of U-Net and its variant techniques for image segmentation and specifically highlighted their applications in retinal fundus image segmentation (36). Yuliang et al. propose a retinal blood vessel segmentation model that combines a multiscale matched filter with a U-Net that has been tested on various available public datasets (37). Shabbir et al. describe different types of ML models for glaucoma detection from fundus image like multiscale, texture feature-based, Segmentation-based, CNN, Ensemble Learning approaches in detail (38). Wang et al. proposed a structure using Dense U-net and the patch-based learning approach for clinical applications (39). **Table 1** given below provides a tabular

representation of the related work discussed in the literature review section.

METHODOLOGY

The end-to-end implementation of retinal vessel detection requires introducing a correct set of features per pixel in which the model can find meaningful discrimination between vessels and background. Concerning the intensity of each pixel in the fundus image, $I(x, y)$ can be defined where x and y are the pixel intensity, and I is the content of the pixel intensity reported in the RGB vector. Several studies have suggested a preprocessing step to enhance the brightness of fundus images (29, 40).

Pre-processing

Illustrated in **Figure 1**, the original fundus images present low-intensity differences among the retina tissue and vessels; therefore, it is proposed to use contrast limited adaptive histogram equalization operation (known as CLAHE) to improve the overall image distinction (1). The CLAHE operation was performed for each color channel to normalize the pixel intensity values image *via* (1).

$$I\left(x,y\right) = \frac{I\left(x,y\right) - u}{\sigma} \tag{1}$$

Where μ is mean, and σ is the standard deviation of each channel.

Concerning the red nature of blood vessels, since the retina texture is supplied *via* micro-vessels, the mean intensity of the red channel is mostly higher than other RGB channels. To this end, it has been reported to remove the red channel from the processing of retinal vessels. In contrast, the green channel and the blue one can be used instead to provide more information about the spread of blood vessels. To this end, an initial step comprising the decomposition of the channel is required, and the omitted channel will be replaced with an extra feature channel to highlight the presence of the vessels across the retina texture.

Construction of the Feature Map

As discussed through the literature review, the cross-section can be modeled as a Gaussian curve; however, in (10), it was shown that Cauchy PDF (2) could be fitted better than the traditional Gaussian in case, as they are compared in **Figure 2**.

$$C\left(x\right) = \frac{1}{\pi\gamma\left[1 + \left(\frac{(x-x_0)}{\gamma}\right)^2\right]} \tag{2}$$

Where γ and x stand for scaling parameter (defining the slope of the curve) and pixel intensity, respectively. By selecting non-zero values for x_0, the curve peak is shifted horizontally. To generate the bank of filters, γ is chosen as 1, the template curves trunk value (T) is selected 9, while the length value of each template (L) was chosen 8, upon the recommendations for DRIVE dataset in (2).

The template is then rotated every $\frac{\pi}{12}$ to detect vessels in different orientations. To prepare an image for either training or testing, each pixel is first convolved with the bank of filters, and the highest response is stored. **Figure 3** depicts an image with its response after the convolution.

As the red channel was initially removed from the original image, the matched filter response is inserted in the red channel. To prepare the new feature images (Cauchy-G-B) for creating patches. Since there are few training images (20 for the dataset), a total of 500 overlapped 64 × 64 patches have been generated per training image. Besides, data augmentation has been employed via either flipping patches horizontal and vertically. This operation assured that more than 40k patches are available for the training and validation process. However, for testing purposes, the total ROI of each test image is segmented into 64 × 64 patches with no overlap. It assures that as soon as there is a prediction per each patch, they can be regrouped structurally to form the predicted image entirely. The image preparation pipeline and the patches are presented in **Figure 4**.

U-Net

The U-Net is widely employed among the different CNN architectures used for retinal vessel detection (41). U-net has been proven to have a promising output in various segmentation approaches, especially in retinal vessel segmentation (3, 4). Concerning the end-to-end training of pixels, it uses multi-level decomposition to incorporate low-level features and multi-channel filtering. On the one hand, employing U-net enhances training useful low-level features. Yet, on the other hand, through the testing, it discards the trained feature, which results in degradation of learning ability.

The employed U-net model is based on a fully connected network structure comprising an input layer, a convolutional layer, a pooling layer, and an up-sampling layer followed by the output layer. Bach to the encoding stage includes four down-sampling steps where there are two convolutional layers and one pooling layer in each. The encoding stage consists of four up-sampling steps, while each one consists of two convolutional layers and one up-sampling layer. The input layer is set to accept patches of the size of 64 × 64 × 3, while the output is designed to generate a 64 × 64 × 1 floating point image. The architecture consists of 32 filters with padding for the size equal to the input for the first and second convolutional layers in the internal layers. The Rectified Linear Unit (ReLU) activation function was used in the whole model except the last layer where SoftMax was employed presented in (3).

$$p_k\left(x\right) = \frac{exp\left[a_k\left(x,y\right)\right]}{\sum_{k'=1}^{K} exp\left[a_k\left(x,y\right)\right]} \tag{3}$$

where $a_k\left(x\right)$ stands for the activation in feature channel k at the pixel position of $\left(x,y\right)$. The number of classes is shown by K and $p_k\left(x\right)$ shows the approximated maximum function. The U-net structure is depicted in **Figure 5**.

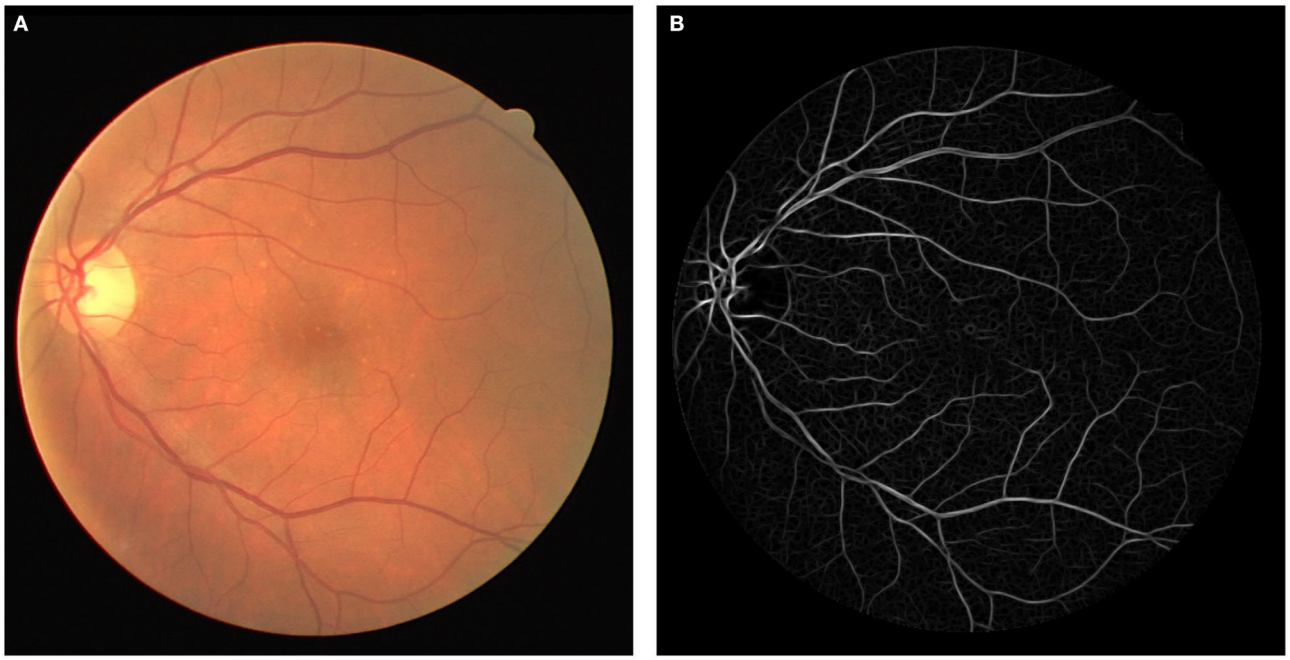


FIGURE 3 | (A) Original fundus image and **(B)** Cauchy matched filter response.

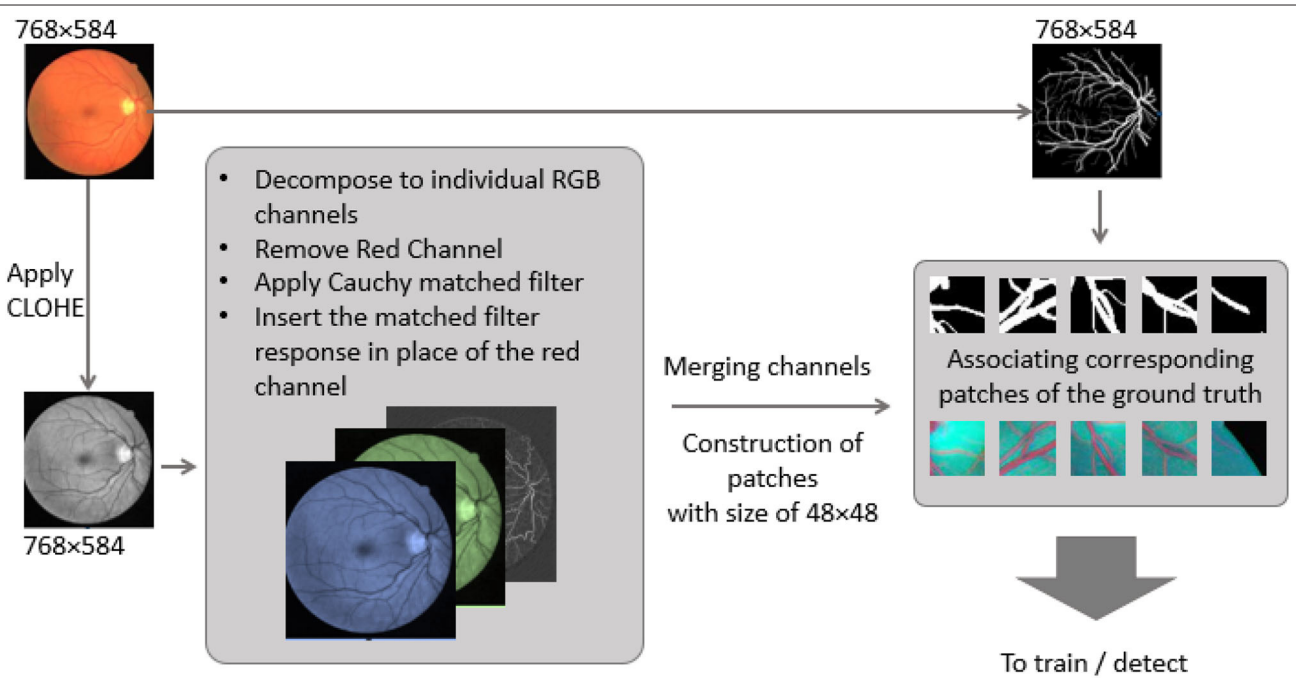


FIGURE 4 | Image preparation pipeline, from original Fundus image to feature map patches.

RESULTS AND DISCUSSIONS

Database

The DRIVE database contains 40 pictures with 768 × 584 pixels captured digitally via a Canon CR5 camera at 45 fields of view (42). Total seven images show some lesions corresponding to degrees of retinopathy. Besides, binary maps containing manual segmentation of vessels are provided, and masks cover retina boundaries. Unfortunately, the images are provided in a lossy compression format (JPEG), which has affected some small

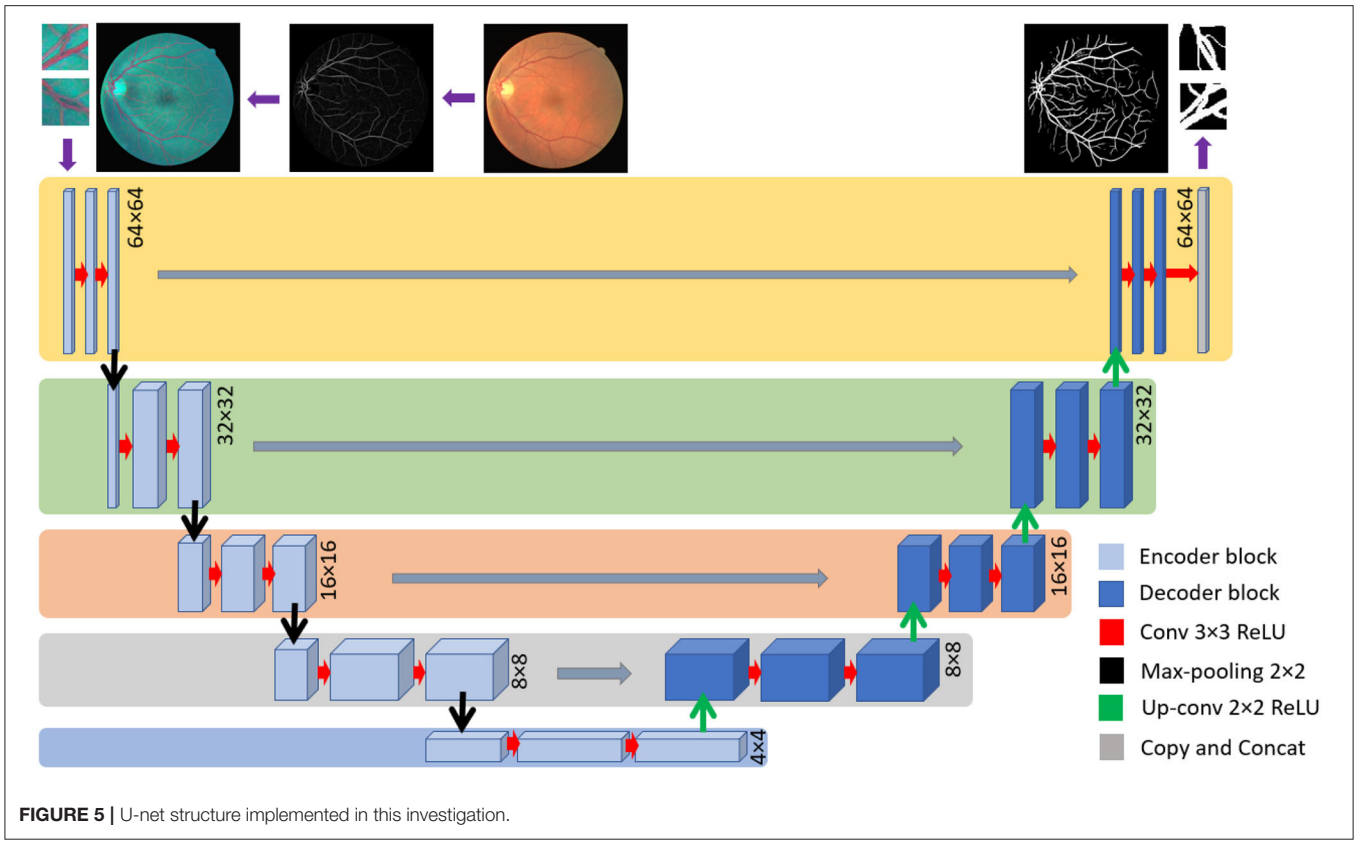


FIGURE 5 | U-net structure implemented in this investigation.

details; however, most researchers have used this dataset as a reference for evaluation. The images are split into two groups, training and test, each containing 20 images with three images presenting retinopathy symptoms.

Implementation Details

We first extracted patches from DRIVE retinal images to train a DL model. Later, we fed them as the input of the designed network and the corresponding label patches obtained from ground truth data. In other words, With the approach used to train a U-net with the end-to-end strategy, the preprocessed patches and their corresponding ground truth extractions were employed to obtain a specific training model for DRIVE database images. Data splitting was carried out by considering 90% of all extracted patches for training and 10% for validation. This strategy has been adopted in several previous studies, including in (42–44), and (33). A total number of 200 epochs were considered for the training process, as well.

Training Details

Adam optimizer was employed to speed up the training process and avoid trapping on local minima while training. It included setting $\beta_2 = 0.9$ and $\beta_2 = 0.999$ to handle sparse gradients on noisy (45). While the initial learning rate was set to $\eta = 0.01$, and it was set to decrease with the shrinking rate of $\eta_i = 0.9\eta_i - 1$ per each 10 epochs when the value of loss function is saturated.

The system configuration used for the training and testing process includes an Intel Core i7-7300 CPU with 64GB of memory. The code was developed in Python 3.7 using Anaconda and Keras as the DL framework and TensorFlow. The preprocessing, including the preparation of multi-channel feature images, was quite fast per each image (<10 s), but instead, each training epoch required more than 2 h to complete.

Evaluation Metric

Retinal vessel segmentation is commonly looked at as a binary classification task. The predicted labels for each pixel are compared with the ground truth classes to fit vessel (positive) or retinal tissue (negative) categories. Binary classification requires determining a true or false class (vessel or retinal tissue) class per pixel. Therefore, a predicted pixel can be represented in four groups:

- True Positive (T_P) are those pixels predicted as vessels and belong to vessels in ground truth images (correct prediction).
- True Negative (T_N) pixels are predicted as non-vessels and belong to retinal tissue (correct prediction).
- False Positive (F_P) indicates the predicted pixels as retinal background but instead belong to vessels (incorrect prediction).
- Finally, False Negative (F_N) are non-vessel pixels, but they have been classified as retinal vessels (incorrect prediction).

For the evaluation purposes, the performance of the trained models is obtained through measuring Sensitivity (Sen), specificity (Spc), accuracy (Acc), and Precision (Prc), which are

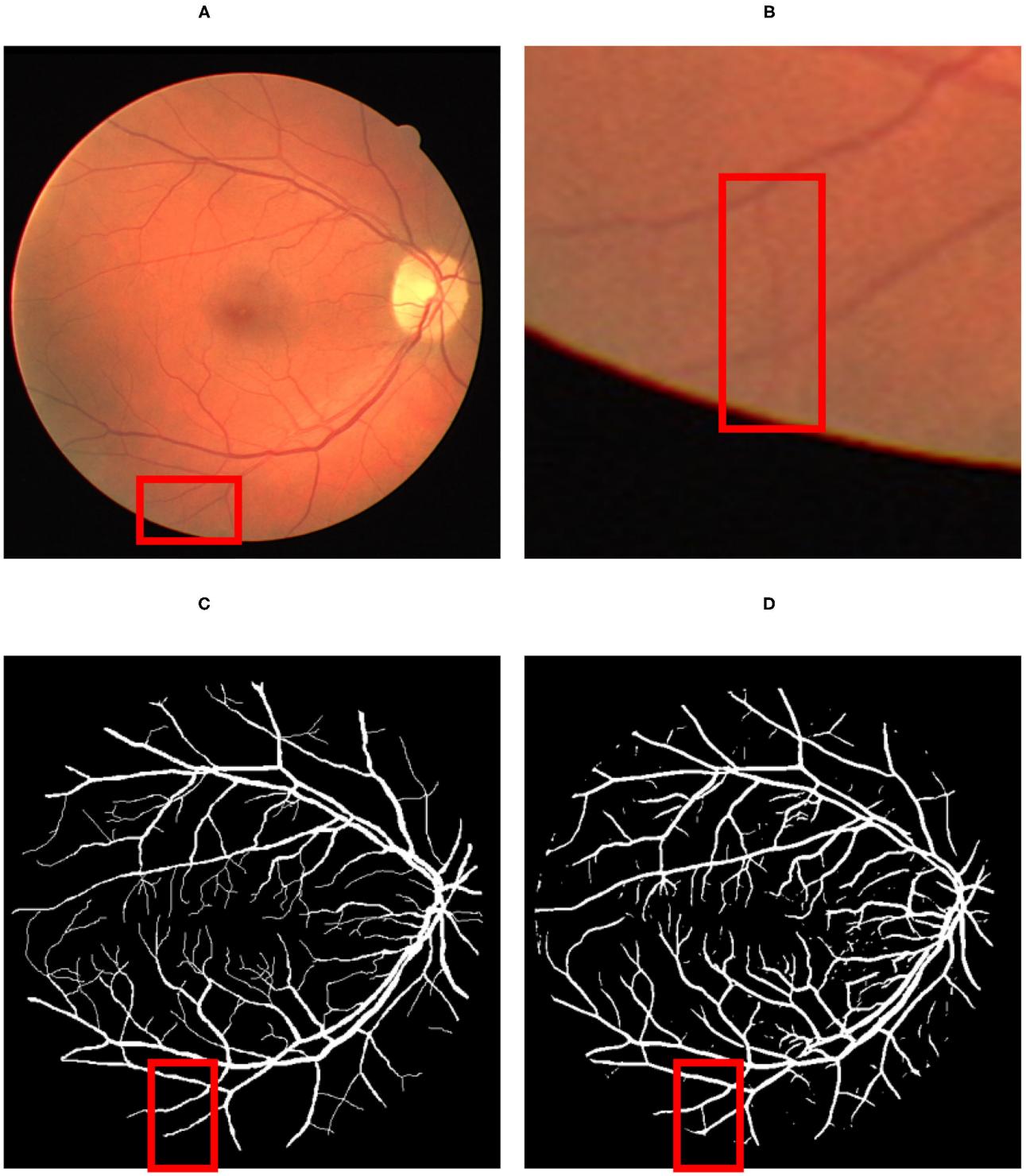


FIGURE 6 | The result of the detection: **(A)** Original fundus image; **(B)** a thin vessel; **(C)** the corresponding ground truth map which does not contain the vessel; **(D)** the thin vessel being detected by the proposed model. The red color square box part in **(A)** has been reflected by respective identified image in **(C)** by proposed map. Similarly the red square part reflected in **(B)** has been identified by red square box part in image **(D)**.

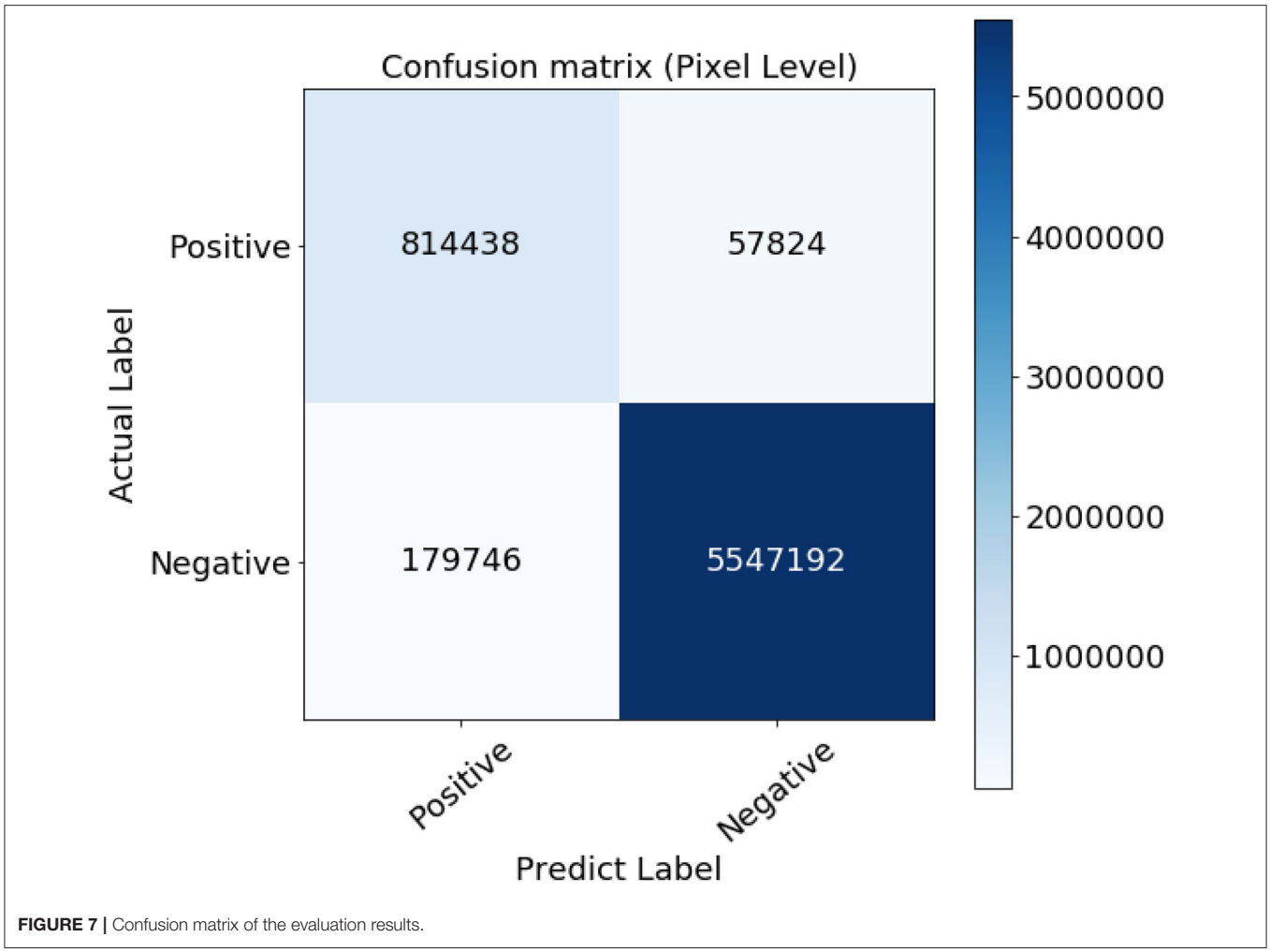


FIGURE 7 | Confusion matrix of the evaluation results.

standard approaches for evaluating binary results represented in (4–7).

$$Acc = \frac{T_P + T_N}{T_p + F_P + T_N + F_N} \tag{4}$$

$$Sen = \frac{T_P}{T_p + F_N} \tag{5}$$

$$Spc = \frac{T_N}{F_P + T_N} \tag{6}$$

$$Prs = \frac{T_P}{F_P + T_P} \tag{7}$$

In addition, the area under the Receiver Operating Characteristic curve (AUROC) for comparison with contemporary methods is obtained, as well.

The overall output of the proposed method is obtained via the receiver operating characteristic curve (ROC), which reflects the relationship between true positive and false positive rates concerning different thresholds when the network output is mapped into binary classes. For this purpose, global thresholding method with different threshold values was considered for the threshold-sensitive metrics by calculating sensitivity against the false-positive ratio ($FPR = \frac{F_P}{F_N + T_P}$) at various threshold values, and later plotting ROC curve per each FPR and thresholding rate pair. The same strategy is adopted for obtaining the average

TABLE 2 | Assessment of the obtained results with contemporary approaches.

Method	Acc	Sen	Spc	Prc	AUROC
Fraz et al. (40)	0.9422	0.7302	0.9742	0.9600	0.9502
Nicola et al. (48)	0.9467	0.7731	0.9724	NA	0.9588
Lahiri et al. (24)	0.9480	0.7500	0.9800	NA	0.9500
Melinscak et al. (27)	0.9466	0.7276	0.9785	NA	0.9749
Fu et al. (47)	0.9470	0.7294	NA	NA	0.9523
Gao et al. (33)	0.9636	0.7802	0.7802	0.8948	0.9771
Mou et al. (43)	0.9607	0.8132	0.9783	NA	0.9796
Fan et al. (44)	0.9664	0.8374	0.9790	NA	0.9835
Wang et al. (39)	0.9511	0.7986	0.9736	NA	0.9740
Saroj et al. (46)	0.9509	0.7278	0.9724	NA	0.8501
Proposed method	**0.9640**	**0.8192**	**0.9896**	**0.9337**	**0.9805**

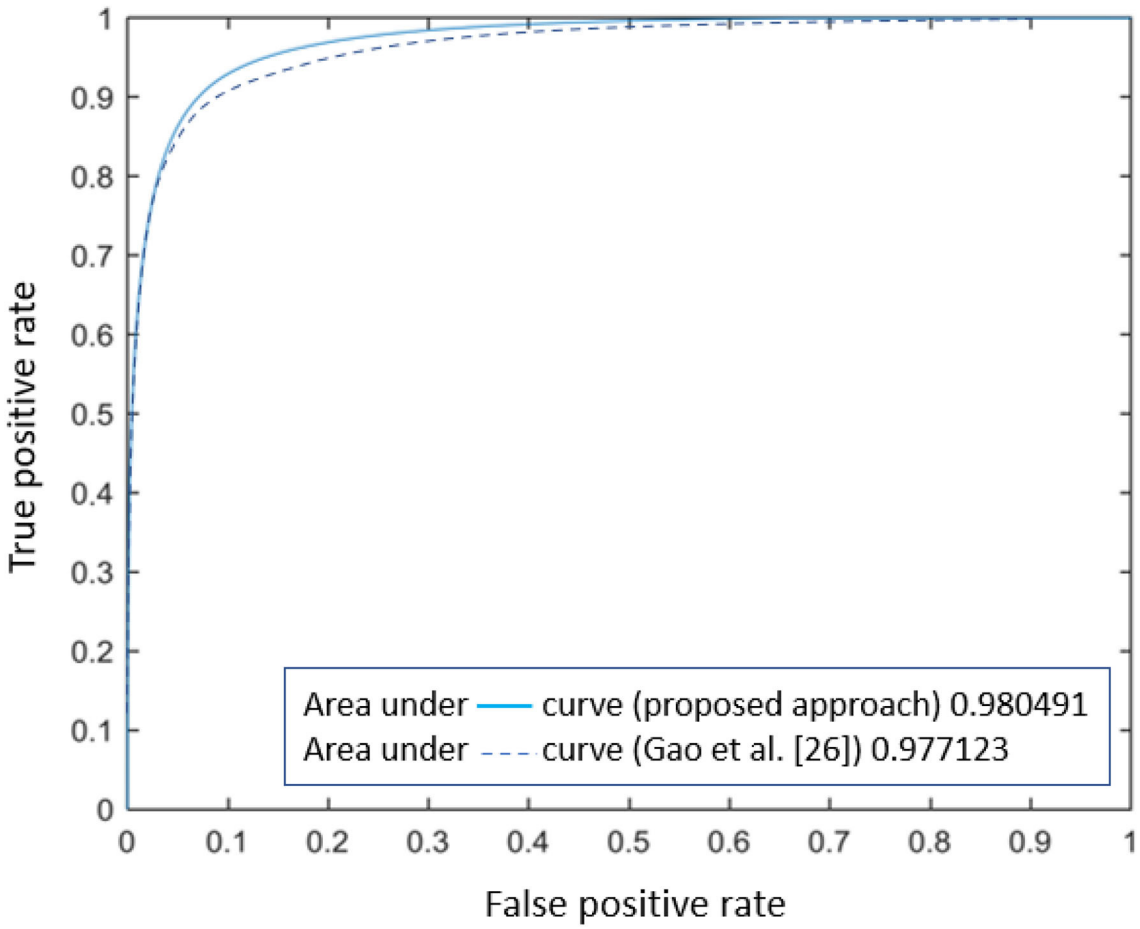


FIGURE 8 | The comparison of ROC curves for retinal vessel detection methods.

precision obtained *via* (7) against the threshold at various threshold values.

Sensitivity Analysis of Global Threshold

To have robust evaluations for metrics that are sensitive to the employed global thresholding method (Acc, Sen, Spc, and Prc), a range of values has been studied threshold values sampled in $\tau_i \in \{0.01, \ldots, 0.99\}$. The study revealed that the changes in the metrics are slight around $\tau = 0.5$, which shows that this threshold value can be considered to be used for further evaluations.

Discussion and Comparison With Previous Approaches

Based on the obtained results, it can be discussed that the inclusion of an additional feature channel (the Cauchy matched filter response) improved the detection rate of vessels via enhancing vessel contrast against the background. In contrast, this strategy filtered the noisy red channel. Besides, vessels with various diameters were segmented by having a bank of filters at different scales for the matched filter. Vessels with multiple diameters were segmented, which assisted the DL method to improve the proposed classifier.

Concerning the subjective characteristic of manual labeling of the ground-truth data, a detailed visual comparison between the predicted vessel map (**Figure 6D**) and the ground-truth (**Figure 6C**) reveals that the proposed method can extract more vessels that had not been labeled before in ground truth images (**Figures 6A,B**).

However, since those new extracted vessels are not included in the subjective extracted ground truth images, total T_P and T_N pixels decreased while F_P and F_N increased, respectively. Such imbalance labeling has impressed the metrics to approach slightly toward a perfect detection. Hopefully, since the ground truth data is common for all the compared methodologies, they can be evaluated simultaneously. The confusion matrix presented in **Figure 7** certifies the details of the evaluation.

Additional visual comparison asserts that the proposed methods have detected thinner vessels in the central regions on the retina, where vessels are distributed in low density. It is due to the vessels' shape to fit better on the employed Cauchy curve.

Numerical comparison with other approaches (reported in **Table 2**) certifies the proposed model is ranked among most state-of-the-art methods, based on Sensitivity, Specificity, accuracy, and area under the ROC. However, some methods are ranked higher [such as the one presented in (44)], as they employed octave convolutions to extract multiple-spatial-frequency features aiming to capture retinal vasculatures with different thicknesses. This technique allowed them to improve the U-net structure and, at the same time, take benefit from features to enhance the detection rate significantly. **Figure 8** illustrates the AUROC of the proposed method, which shows a bigger area under the curve (0.9805) compared to the baseline model (0.9771) presented by Gao et al. (33).

Wang et al. Proposed a model that provides an accuracy rate of 0.9511 (39), while the model proposed by Saroj et al. Shows an accuracy rate of 0.9509 (46). Therefore, it can be concluded that the proposed model's accuracy in this paper is superior to existing models.

CONCLUSION AND FUTURE WORK

Targeting retinal vessel detection, this paper offers a novel approach that pools the response of a traditional matched filter with a fully connected U-net network. The uniqueness of the suggested method is to replace the noisy red channel of fundus images with a channel that can provide more vessel details. It is intended since the red blood vessel are barely visible distributed over the red retina tissue. The new pseudo-color image, called a multi-channel feature map, is processed by generating several patches for training via the U-net. The corresponding labels for the training patches are also provided using the ground truth data. This technique accelerates the training process and provides enough data to train the DL model. Evaluation of the results shows that vessel detection can be accomplished with an accuracy of 96.40% and sensitivity around 82%. This difference is seen since thin vessels have been discarded during the subjective annotation of ground truth images. The blue channel of retinal images can be replaced with a Gabor-filter response in future work. The bank of the used filters can be equipped with vessels presented in a variety of widths. Additional channels are also intended to be investigated to assist the DL model in associating the pixel values to the vessel or retinal tissue classes. Last but not least, and through comparison of this work to the other state-of-the-art methodologies, the proposed method shows a promising result highlighting the importance of combining feature maps to improve the detection of retinal vessels via U-net assisted DL approaches.

Gathering large quantities of fundus images is one of the most perplexing in creating strong deep-learning systems. Designing algorithms that can work from restricted data is one solution. Apart from limited images available in public datasets, there is still an opportunity for testing blood vessel segmentation approaches on noisy real-life pathological retinal pictures. The offered model has been tested only on the DRIVE dataset. Still, it can also be tested on the other available datasets to evaluate its accuracy and compare with other models. Also, the proposed model needs to be tested to real-time pathological retinal images.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SA and MS: conceptualization, writing—original draft, and preparation. SA, MH, and SB: methodology. SB: validation. MH and FJ: resources. SB, MH, and RA: writing—review and editing and funding acquisition. MS, FJ, and SA: supervision. SA, RA, and SB: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Deanship of Scientific Research, King Faisal University, Saudi Arabia (grant number NA000256).

REFERENCES

1. Jusoh F, Haron H, Ibrahim R, Che Azemin MZ. *An Overview of Retinal Blood Vessels Segmentation BT - Advanced Computer and Communication Engineering Technology*. Phuket: Springer International Publishing. (2016) p. 63–71. doi: 10.1007/978-3-319-24584-3_6
2. GeethaRamani R, Balasubramanian L. Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis. *Biocybern Biomed Eng.* (2016) 36:102–18. doi: 10.1016/j.bbe.2015.06.004
3. Fraz MM, Remagnino P, Hoppee A, Uyyanonvara B, Rudnika AR, Owen CG, et al. Blood vessel segmentation methodologies in retinal images-a survey. *Comp Meth Prog Biosci*. (2012) 108:407–33. doi: 10.1016/j.cmpb.2012.03.009
4. Soares J, Leandro J, Cesar Junior R, Jelinek H, Cree M. Retianl vessel segmentation using the 2-D gabor wavelet and supervised classification. *IEEE Trans Med Imaging*. 25:1214–1222. doi: 10.1109/TMI.2006.879967
5. Orujov F, Maskeliunas R, Damaševičius R, Wei W. Fuzzy based image edge detection algorithm for blood vessel detection in retinal images. *Appl Soft Comput*. (2020) 94:106452. doi: 10.1016/j.asoc.2020.106452

6. Leopold HA, Orchard J, Zelek J, Lakshminarayanan V. Use of Gabor filters and deep networks in the segmentation of retinal vessel morphology. In: *Proc. SPIE*. (2017). doi: 10.1117/12.2252988
7. Imran J, Li Y, Pei J, Yang J, Wang Q. Comparative analysis of vessel segmentation techniques in retinal images. *IEEE Access*. (2019) 7:114862–87. doi: 10.1109/ACCESS.2019.2935912
8. Vermeer KA, Vos FM, Lemij HG, Vossepoel AM, A model based method for retinal blood vessel detection. *Comput Biol Med*. (2004) 34:209–219. doi: 10.1016/S0010-4825(03)00055-6
9. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans Med Imaging*. (1989) 8:263–269. doi: 10.1109/42.34715
10. Al-Rawi M, Qutaishat M, Arrar M. An improved matched filter for blood vessel detection of digital retinal images. *Comput Biol Med*. (2007) 37:262–267. doi: 10.1016/j.compbiomed.2006.03.003
11. Wu C, Agam G, Stanchev P. A general framework for vessel segmentation in retinal images. in 2007. I*nternational Symposium on Computational Intelligence in Robotics and Automation*, (2007). 37–42. doi: 10.1109/CIRA.2007.382924
12. Dizdaro B, Ataer-Cansizoglu E, Kalpathy-Cramer J, Keck K, Chiang MF, Erdogmus D. Level sets for retinal vasculature segmentation using seeds from ridges and edges from phase maps. I*EEE International Workshop on Machine Learning for Signal Processing*. (2012) 1–6. doi: 10.1109/MLSP.2012.6349730
13. Zhang J, Tang Z, Gui W, Liu J. Retinal vessel image segmentation based on correlational open active contours model. C*hinese Automation Congress (CAC)*. (2015) 993–8.
14. Ali A, Zaki WMDW, Hussain A. Retinal blood vessel segmentation from retinal image using B-COSFIRE and adaptive thresholding. *Indones J Electr Eng Comput Sci*. (2019) 13:1199–207. doi: 10.11591/ijeecs.v13.i3.pp1199-1207
15. Dhanamjayulu C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) ipr2.12222.
16. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic deep Q network for real-time path planning in censorious robotic procedures using force sensors. *J. Real-Time Image Process*. (2021) 18: 5,1773–1785. doi: 10.1007/s11554-021-01122-x
17. Rehman A, Harouni M, Karimi M, Saba T, Bahaj SA, Awan MJ. Microscopic retinal blood vessels detection and segmentation using support vector machine and K-nearest neighbors. *Microsc Res Tech*. (2022) 1–16. doi: 10.1002/jemt.24051
18. Gu SH, Nicolas V, Lalis A, Sathirapongsasuti N, Yanagihara R. Complete genome sequence and molecular phylogeny of a newfound hantavirus harbored by the Doucet's musk shrew (Crocidura douceti) in Guinea. *Infect Genet Evol*. (2021) 20:118–123 doi: 10.1016/j.meegid.2013.08.016
19. Singla M, Soni S, Saini P, Chaudhary A, Shukla KK. Diabetic retinopathy detection using twin support vector machines. *Adv Intell Syst Comput*. (2020) 1064:91–104. doi: 10.1007/978-981-15-0339-9_9
20. Badar M, Haris M, Fathima A. Application of deep learning for retinal image analysis: A review. *Comp Sci Rev*. (2020) 35:100203. doi: 10.1016/j.cosrev.2019.100203
21. Gadekallu TR, Rajput DS, Reddy MPK, Lakshmanna K, Bhattacharya S, Singh S. A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *J Real Time Image Process*. (2020) 18:1383–96. doi: 10.1007/s11554-020-00987-8
22. Wu H, Zhao S, Zhang X, Sang A, Dong J, Jiang K. Back-propagation artificial neural network for early diabetic retinopathy detection based on a priori knowledge. *J Phys: Conf Serv*. (2020) 1437:012019. doi: 10.1088/1742-6596/1437/1/012019
23. Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*. (2015) 149:708–717. doi: 10.1016/j.neucom.2014.07.059
24. Lahiri A, Roy AG, Sheet D, Biswas PK. Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. 3*8th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. (2016) 1340–1343, doi: 10.1109/EMBC.2016.7590955
25. Maji D, Santara A, Mitra P, Sheet D. Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images. *CoRR*. (2016) ArXiv abs/1603.04833:1–4.
26. Liskowski P, Krawiec K. Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging*. (2016) 35:2369–20. doi: 10.1109/TMI.2016.2546227
27. Melinscak M, Prentasic P, Loncaric S. Retinal vessel segmentation using deep neural networks. *VISAPP* 2015. - *10th Int Conf Comput Vis Theory Appl VISIGRAPP, Proc*. (2015) 1:577–582. doi: 10.5220/0005313005770582
28. Dasgupta A, Singh S. A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. I*EEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. (2017) 248–251. doi: 10.1109/ISBI.2017.7950512
29. Wu A, Xu Z, Gao M, Buty M, Mollura DJ. *Deep Vessel Tracking: A Generalized Probabilistic Approach via Deep Learning*. (2016). Available online at: ieeexplore.ieee.org doi: 10.1109/ISBI.2016.7493520
30. Fu H, Xu Y, Lin S, Wong DWK, Liu J. Deepvessel: retinal vessel segmentation *via* deep learning and conditional random field. *Lect Notes Comput. Sci*. (2016) 9901:132–139 doi: 10.1007/978-3-319-46723-8_16
31. Quasim MT, Shaikh A, Shuaib M, Sulaiman A, Alam S, Asiri Y. Smart healthcare management evaluation using fuzzy decision making method. *Res Sq*. (2021) 1–19. doi: 10.21203/rs.3.rs-424702/v1
32. Upreti K, Syed M, Alam M, Alhudhaif A, Shuaib M. Generative adversarial networks based cognitive feedback analytics system for integrated cyber-physical system and industrial iot networks. *Res Sq*. (2021) 1–14. doi: 10.21203/rs.3.rs-924288/v1
33. Gao X, Cai Y, Qiu C, Cui Y. Retinal blood vessel segmentation based on the Gaussian matched filter and U-net. 1*0th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. (2017) 1–5. doi: 10.1109/CISP-BMEI.2017.8302199
34. Zolfagharnasab H, Naghsh-Nilchi AR. Cauchy based matched filter for retinal vessels detection. *J Med Signals Sens*. (2014) 4:1. doi: 10.4103/2228-7477.128432
35. Roy A, KS. *Retinal vessel detection using residual Y-net*. (2021). Available online at: ieeexplore.ieee.org doi: 10.1109/SMARTGENCON51891.2021.9645749
36. Siddique N, Paheding S. Access E-I, undefined. *U-net and its variants for medical image segmentation: A review of theory and applications*. (2021). Available online at: ieeexplore.ieee.org doi: 10.1109/ACCESS.2021.3086020
37. Ma Y, Zhu Z, Dong Z, Shen T. *Multichannel Retinal Blood Vessel Segmentation Based on the Combination of Matched Filter and U-Net Network*. (2021). Available online at: hindawi.com doi: 10.1155/2021/5561125
38. Shabbir A, Rasheed A, Shehraz H, Saleem A, Zafar B, Sajid M, et al. Blood vessel segmentation methodologies in retinal images-a survey. *Math Biosci Eng*. (2021) 18:2033–76. doi: 10.3934/mbe.2021106
39. Wang C, Zhao Z, Ren Q, Xu Y. Dense U-net based on patch-based learning for retinal vessel segmentation. (2019). Available online at: mdpi.com doi: 10.3390/e21020168
40. Fraz MM, Basit A, Barman SA. Application of morphological bit planes in retinal blood vessel extraction. *J Digit Imaging*. (2013) 26:274–286. doi: 10.1007/s10278-012-9513-3
41. Gadekallu TR, et al. Hand gesture classification using a novel CNN-crow search algorithm. *Complex Intell Syst*. (2021) 7:1855–68. doi: 10.1007/s40747-021-00324-x
42. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging*. (2004) 23:501–9. doi: 10.1109/TMI.2004.825627
43. Mou L, Chen L, Cheng J, Gu Z, Zhao Y, Liu J. Dense dilated network with probability regularized walk for vessel detection. *IEEE Trans Med Imaging*. (2020) 39:1392–403. doi: 10.1109/TMI.2019.2950051
44. Fan Z, et al. Accurate retinal vessel segmentation *via* octave convolution neural network. *arXiv*. (2019) abs/1906.1.
45. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. *Proc IEEE Int Conf Comput Vis*. (2015) 2015:1026–1034. doi: 10.1109/ICCV.2015.123
46. Saroj S, Kumar R. *Frechet PDF based matched filter approach for retinal blood vessels segmentation*. Elsevier. (2020). doi: 10.1016/j.cmpb.2020.105490
47. Fu H, Xu Y, Wong DWK, Liu J. Retinal vessel segmentation *via* deep learning network and fully-connected conditional random fields. In: *Proceedings - International Symposium on Biomedical Imaging*. (2016) p. 698–701. doi: 10.1109/ISBI.2016.7493362

48. Strisciuglio N, Azzopardi G, Vento M, Petkov N. Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters. *Mach Vis Appl.* (2016) 27:1137–49. doi: 10.1007/s00138-016-0781-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bhatia, Alam, Shuaib, Hameed Alhameed, Jeribi and Alsuwailem. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fpubh.2022.860536

Classification of Non-Functional Requirements From IoT Oriented Healthcare Requirement Document

Iqra Khurshid[1], Salma Imtiaz[1], Wadii Boulila[2], Zahid Khan[2], Almas Abbasi[1], Abdul Rehman Javed[3]* and Zunera Jalil[3]*

[1] Department of Software Engineering, International Islamic University, Islamabad, Pakistan, [2] Robotics and Internet-of-Things Laboratory, Prince Sultan University, Riyadh, Saudi Arabia, [3] Department of Cyber Security, Air University, Islamabad, Pakistan

Internet of Things (IoT) involves a set of devices that aids in achieving a smart environment. Healthcare systems, which are IoT-oriented, provide monitoring services of patients' data and help take immediate steps in an emergency. Currently, machine learning-based techniques are adopted to ensure security and other non-functional requirements in smart health care systems. However, no attention is given to classifying the non-functional requirements from requirement documents. The manual process of classifying the non-functional requirements from documents is erroneous and laborious. Missing non-functional requirements in the Requirement Engineering (RE) phase results in IoT oriented healthcare system with compromised security and performance. In this research, an experiment is performed where non-functional requirements are classified from the IoT-oriented healthcare system's requirement document. The machine learning algorithms considered for classification are Logistic Regression (LR), Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), K-Nearest Neighbors (KNN), ensemble, Random Forest (RF), and hybrid KNN rule-based machine learning (ML) algorithms. The results show that our novel hybrid KNN rule-based machine learning algorithm outperforms others by showing an average classification accuracy of 75.9% in classifying non-functional requirements from IoT-oriented healthcare requirement documents. This research is not only novel in its concept of using a machine learning approach for classification of non-functional requirements from IoT-oriented healthcare system requirement documents, but it also proposes a novel hybrid KNN-rule based machine learning algorithm for classification with better accuracy. A new dataset is also created for classification purposes, comprising requirements related to IoT-oriented healthcare systems. However, since this dataset is small and consists of only 104 requirements, this might affect the generalizability of the results of this research.

Keywords: non-functional requirements, healthcare, classification, machine learning, requirement document

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Prabhat Kumar, Indian Institute of Technology Hyderabad, India
Sharnil Pandya, Symbiosis International University, India

***Correspondence:**
Wadii Boulila
wboulila@psu.edu.sa
Abdul Rehman Javed
abdulrehman.cs@au.edu.pk

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *23 January 2022*
Accepted: *07 February 2022*
Published: *18 March 2022*

Citation:
Khurshid I, Imtiaz S, Boulila W, Khan Z, Abbasi A, Javed AR and Jalil Z (2022) Classification of Non-Functional Requirements From IoT Oriented Healthcare Requirement Document. Front. Public Health 10:860536. doi: 10.3389/fpubh.2022.860536

1. INTRODUCTION

One of the most important tasks of developing high-quality software is gathering the right requirements and ensuring no missing requirements. Often during the process of Requirement Engineering (RE), more attention is given to eliciting the functional requirements than non-functional requirements. This results in the poor quality end product and results in loss of

cost, effort, and even failure of the project. Non-functional requirements explain the important quality attributes (1, 2) and constraints (3) that must be implemented in the system. If these requirements are left ignored, the architecture is not designed properly (4). Thus, classifying the non-functional requirements becomes a mandatory task in the RE process for designing the architecture design and performing other related activities accordingly (4, 5). Traditionally, non-functional requirements are identified and extracted manually which is an erroneous process (5, 6) and laborious (7). The non-functional requirements are intertwined with the functional requirements in RE documents, not properly structured. Therefore, proper identification and categorization are required (8). Since the architect needs to know the types of non-functional requirements, non-functional requirements must be classified properly (9). Currently, the trend of using ML algorithms for classifying the non-functional requirements is gaining attention (8). However, there is still a need to classify the non-functional requirements accurately. By using automated techniques, the chance of missing the non-functional requirements is reduced, classification accuracy can be improved, and time and effort are also saved. IoT possesses a vital role in the health care systems of the present era. It includes a set of devices, which help store, process, and transfer data to achieve smart services (10–13). Healthcare systems, which are IoT-oriented, provide monitoring services of patients' data and help take immediate steps in an emergency (14–16). Thus, it is critical to ensure that these smart healthcare systems are extremely secure and perform reliably so that patients' sensitive data is not only kept protected from manipulation and attacks but also transferred entirely and safely in a timely fashion (17–19). In order to ensure that the IoT-oriented healthcare system possesses all these important non-functional requirements, it is necessary to extract all the non-functional requirements from the requirement document of the IoT-oriented healthcare system in the RE phase. Machine learning-based techniques are adopted to implement non-functional requirements in smart health care systems. However, attention needs to be given to the first classification of the non-functional requirements from the requirement document to ensure that non-functional requirements are not missed in the RE phase. By performing this classification task automatically, the chances of missing non-functional requirements will be reduced, and the probability of developing a high-performing and secure IoT-oriented healthcare system will increase (20).

In this research, the main focus is on finding machine learning algorithm and relevant features which helps in the classification of non-functional requirements with higher accuracy. For features extraction, Bag of Words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF) are adopted. The algorithms considered by this study are Logistic Regression (LR), Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), K-Nearest Neighbors (KNN), ensemble, Random Forest (RF), and hybrid KNN rule-based ML algorithms. The two main distinct additions here are the ensemble and hybrid classifiers.

The existing studies have limitations. Some studies have reported low performance (1, 6, 21). The types of non-functional requirements considered by some studies are very less (4, 22, 23). The dataset adopted by different studies has a limited number of non-functional requirements (2, 9, 21). Some studies adopted the manual process of validation which may have chances of error (21).

This article makes the following contributions.

- Creation of dataset, which includes requirements related to IoT-oriented healthcare system. The requirements included in this dataset belong to 8 categories: Accuracy, Reliability, Security and Privacy, Performance, Compatibility, Usability, Functional, and Maintainability. The numbers of requirements in this dataset are 104. IoT-oriented healthcare system requirement documents create this dataset (24, 25). Since there are no currently such dataset available that contains labeled requirements related to IoT-oriented healthcare systems, this will serve as an aid for future research purposes in automating the classification of non-functional requirements from IoT-oriented healthcare system requirement documents.
- Development of a novel hybrid KNN rule-based machine learning algorithm, which provides better classification accuracy than traditional machine learning algorithms like SVM, KNN, RF, MNB, LR, and ensemble. This novel hybrid machine learning algorithm helps classify non-functional requirements from IoT-oriented healthcare system requirement documents with an average classification accuracy of 75.9%.
- Provision of features relevant to non-functional requirements that the IoT-oriented healthcare system must possess. These features will help the researchers to create better classifiers by including them in machine learning-based techniques.

This research article is structured in the following manner. Section 2 describes the relevant work conducted earlier in the literature related to the classification of the non-functional requirements. Section 3 describes the research methodology which is adopted in this research. Section 4 shows the experiment portion. Section 5 presents the results and discussion. Section 6 provides the conclusion and highlights possible future directions.

2. LITERATURE REVIEW

In literature, many studies have highlighted the problem of ignored and missed non-functional requirements in the RE process (8, 21). Non-functional requirements are sometimes hidden and not clearly stated, due to which it becomes difficult to identify them, and chances of missing them increase. Initially, more importance is given to eliciting the functional requirements where non-functional requirements are discovered at late stages which results in issues related to architectural design (1–3, 5, 6), cost management (9, 22, 23), time management (4, 5, 7, 23), and risk (26) and quality management (5, 7, 27, 28). The problem extrapolates due to manual identification and classification of non-functional requirements (5). The manual classification is quite a time-consuming and laborious process (22, 29). Search function provided by tools is used by typing keywords and searching, which is an exhaustive process and

TABLE 1 | Non-functional requirements classification review.

Reference	Machine learning algorithm	Dataset	Limitations
Shreda and Hanani (30)	Convolutional Neural Network	PURE dataset sample	Only 5 non-functional requirement are considered, Probability of biased result due to 2 fold cross validation
Dias Canedo and Cordeiro Mendes (7)	Logistic Regression	PROMISE_exp	Classification error of 25% on average is reported which needs to be minimized
Younas et al. (32)	K Nearest Neighbor	PROMISE	Precision is found as 50.65%, Recall as 41.11% which indicates quite low classification performance.
Lu and Liang (29)	Bagging	21969 user review sentences	Only four non-functional requirement are considered
Kurtanovic and Maalej (33)	Support Vector Machines	PROMISE	Precision is found to be 78.25% which needs be further improved
Maiti (27)	K Nearest Neighbors	EU Procurement documents, NFRM data set, PROMISE	Datasets used for validation contains only 56 and 78 requirements which are very limited Validation is done only on EU procurement documents
Riaz et al. (31)	K Nearest Neighbors	CCHIT, ED, NU, OSCAR	Validation is done specifically on health care domain documents Only security related requirement are considered
Knauss and Ott (26)	Support Vector Machines	2,000 requirements in Mercedes-Benz specifications	Non-Functional requirement types are not mentioned Validation is done only on automotive industry domain document
Rashwan et al. (23)	Support Vector Machines	Corpus with 3064 annotated sentences	Only 5 non-functional requirement are considered Chances of biased results due to 6 cross fold validation
Slankas and Williams (9)	Support Vector Machines	CCHIT Ambulatory Requirements, iTrust, PROMISE	Performance is low in terms of precision and recall which are only 72.8 and 54.4%, respectively
Casamayor et al. (4)	Expectation Maximization with Naïve Bayes	PROMISE	Only accuracy of 75% is reported
Hussain et al. (22)	Decision Trees	PROMISE	Non-Functional requirement types are not stated
Cleland-Huang et al. (5)	Naïve Bayes	PROMISE, SIEMENS IET Datasets	Classification precision is 12.4% which is very low

may lead to missing the non-functional requirement (30). The concept of automating or semi-automating this task gained popularity (7, 27). However, this trend of automating the non-functional requirement classification is still young and requires further research for preprocessing, selecting optimum feature sets, acquiring relevant datasets, and using appropriate ML algorithms (8). Moreover, different studies have worked on different categories of non-functional requirements (8) and they considered different types of documents like Certification Commission for Healthcare Information Technology (CCHIT) Ambulatory requirement document (9), Emergency Department Information Systems Functional Document (31), Mercedes-Benz Specification Document (26), and European Union e-Procurement documents (27) from where the non-functional requirements are extracted. Due to different factors, the classification performance reported by different studies is different, and the comparison is not possible because of the element of biasness. Few studies have managed to report the full process of classifying the non-functional requirements from the requirement document (8). **Table 1** shows various techniques which are adopted by different studies for the classification of non-functional requirements.

2.1. Non-Functional Requirements Classification Using Naive Bayes

One of the techniques to classify the non-functional requirements from the requirement document is by Naive Bayes (5). In this technique, the probability is calculated to classify non-functional requirements. Nine non-functional requirements are focused on, and the PROMISE dataset is used. For evaluation, leave one out cross-validation is performed having 15 iterations. The classifier's performance is very poor precision, which is 12.4%. The classification error of the above classifier indicates that it needs improvement.

Similarly, in another study, the Naive Bayes algorithm is used to classify the non-functional requirement but with Expectation-Maximization (4). Non-functional requirements categories in which the data is classified are 9. The classifier's performance in terms of accuracy is reported to be about 75%.

2.2. Non-Functional Requirements Classification Using Decision Trees

Decision trees are also used to classify the non-functional requirements (22). PROMISE dataset is used, and 10-fold cross-validation is performed for evaluation. The accuracy of classification is 98.56%. The study above has not mentioned the types of non-functional requirements.

2.3. Non-Functional Requirements Classification Using Bagging

Research classifies the non-functional requirements from the requirement document by using machine learning ensemble meta-algorithm Bagging (29). The dataset consisted of 21,969 user reviews. The evaluation is performed using 10-fold cross-validation. The classifier's performance in terms of precision is found to be 71.4% and recall of 72.3%. The non-functional

requirements types that are considered by the study are 4. First of all, the types of non-functional requirements considered are only four. Furthermore, the performance of the classifier is relatively low.

2.4. Non-Functional Requirements Classification Using SVM

In one research, a SVM is used to extract the non-functional requirements from the requirement document (23). In the technique, the documents are first preprocessed. After preprocessing, SVM is applied. Cross-validation is applied for validation in which the dataset is divided into subsets, and validation is performed in iterations. One subset is used for validation in each iteration, and the remaining subsets are used for training. The performance of the classifier is evaluated using 6-fold cross-validation. The performance on the PROMISE corpus in terms of precision is 77%. The non-functional requirements types that are considered are 5. In terms of performance, there is still work that needs to be done to increase the precision. The Sequential Minimal Optimization (SMO) algorithm is used for classifying the non-functional requirements from the requirement document (9). First, preprocessing is done on text, then SMO is applied. The evaluation is performed using 10-fold cross-validation. The performance is found to be 72.8% in terms of precision. The precision of the classifier needs improvement.

Another study reported using the SVM algorithm to classify the non-functional requirements from the specification of Mercedes Benz (26). This approach is semi-supervised. Manual evaluation is performed. The performance of the classifier in terms of precision is >60%. The research is done using an automotive industry document; there is no surety about how the classifier performs in other industry specifications. SVM is used by one more study to extract the non-functional requirements (33). In this technique, the documents are first preprocessed, then applied SVM. The performance of the classifier is evaluated using 10-fold cross-validation. PROMISE dataset is used for classification. The non-functional requirements which are considered by the study are only 4. The performance in terms of precision is found to be 78.25%.

2.5. Non-Functional Requirements Classification Using KNNs

One of the studies focused on classifying security-related requirements (31, 34, 35). In the study, first preprocessing is done, then KNN is applied. The types of security requirements into which the sentences are classified are 7. The evaluation is performed using 10-fold cross-validation. The classifier's performance in terms of precision is found to be 82%. The above study focused on only security-related requirements, not other non-functional requirements.

In literature, another study used KNN to classify the non-functional requirements (27). In the article, first, the text is preprocessed, then KNN is applied. The classifier's performance is found to be 97.73% in terms of precision. The datasets used for evaluation contain requirement documents with 57 and 88 non-functional requirement sentences. The data set used for validation has a minimal number of non-functional requirements.

One of the research used the KNN algorithm concept to classify the non-functional requirements (32). The dataset used for classification is the PROMISE dataset. The classifier's performance is evaluated by iterating 14 times using different threshold values. The classifier's performance in terms of precision is found to be only 50.65%.

2.6. Non-Functional Requirements Classification Using LR

Logistic regression is used by research to classify non-functional requirements (7). This algorithm works based on the probability function. The data set used the PROMISE_exp dataset. The types of non-functional requirements considered for classification purposes are 11. The classifier's performance is calculated by performing 10-fold cross-validation, only 75% in terms of precision.

2.7. Non-Functional Requirements Classification Using Convolutional Neural Network (CNN)

One study proposed CNN for classifying the non-functional requirements from the requirement document (30). We focus on a total of 5 non-functional requirements. The text is first preprocessed in the study, then CNN is applied. A PURE dataset which consists of 1,247 requirement sentences, is used. The classifier's performance is calculated by performing 2-fold cross-validation, and it is found to be 92.2% in terms of precision.

Similarly, various techniques are proposed to ensure the security of IoT-oriented health care systems. One technique proposed is based on a machine-learning algorithm that uses a biometric framework for ensuring the security of sensitive data (36). In this technique, electrocardiogram (ECG) signals are used to extract the features used by a machine-learning algorithm. The proposed system is based on a multilayer perception model, where in order to secure the signal and protect it from possible interference, a secure communication channel is developed. The authentication process is carried out in the testing phase by considering unique generated EIs from the ECG and coefficients from polynomial approximation.

One study highlights the usage of the biometric cryptosystem (BCS) for ensuring security (37). The proposed approach considers Artificial Neural networks (ANN) to analyze the signal energy variations of implanted devices. The inertial measurement units (IMU) are fixed inside the implanted devices, which help detect signal energy changes. In this technique, ANN is trained so that the first sensor is placed on the chest, and then signals are processed by the algorithm. ANN in the proposed technique comprises three layers: an input layer, an output layer, and a hidden layer comprising 10 hidden nodes. The security is achieved by a 128-bit key generated by gait cycles. This key is difficult for hackers since it cannot be achieved by a modern attacking mechanism like dictionary attacks.

In research study, a novel machine learning-based security framework is proposed to detect malicious activities in healthcare systems (38). The proposed framework named Health Guard is developed by considering four machine learning algorithms: KNN, RF, Decision Trees, and ANN. The malicious activities are detected by observing the vital signals of the implanted devices and then correlating the vitals to identify the variations in the patient body. The classifiers are developed by training using nine databases of eight smart implantable devices and by considering activities, among which 7 are normal user activities, and five are disease-related. The validation process is conducted by testing through only three threats: denial of Service (DoS) attacks, tempered medical devices, and false data injection.

The literature review reflects the limitations of the current study. The performance of some techniques is quite average (1, 6, 26). The non-functional requirements classes considered by some studies are very few (4, 23, 31). The dataset contains limited non-functional requirement sentences for training and testing purposes (2, 3, 9). Studies have mostly considered classifiers, which are developed using single machine learning algorithms (8). Different combinations of hybrid and ensemble classifiers are barely considered.

3. RESEARCH METHODOLOGY

In order to find the relevant features and determine the ML algorithm that achieves better accuracy for the classification of non-functional requirements, the research answers the following research questions:

- **RQ.1.** Which features help to increase the accuracy of non-functional requirement classification?
- **RQ.2.** Which machine learning algorithm classifies non-functional requirements with better accuracy?

The research method used for the validation of the ML algorithm is experimentation (39). Experimentation is a systematic method that involves observation, manipulation, and control to generate accurate and reliable results. It is used to study the cause-effect relationship and prefers standardizing tools for maintaining control and achieving precision. It helps develop automatic techniques and evaluate the accuracy of classifying the non-functional requirements while controlling variables like data used for classification, number of non-functional requirements to be classified, and validation method adopted. Simulation, which is similar to experimentation, is not adopted because of the unavailability of its immense requirements to simulate the technique correctly and because of more chances of uncertainty in results due to its abstract nature (40).

PROMISE_exp dataset is used for classification (7). This dataset contains requirements from real projects, and it is the expanded form of the existing PROMISE dataset. The use of a large dataset also helps in generalizing the results (7). Eleven non-functional requirements are chosen for classification: Availability, Legal and Licensing, Look and Feel, Maintainability, Operability, Performance, Scalability, Security, Usability, Fault Tolerance, and Portability. These non-functional requirements are the common non-functional requirements in many projects and are also misclassified and ignored due to their ambiguous nature (41). The ambiguity level is reduced in this research by using features relevant to non-functional requirements. In addition to this, the PROMISE_exp dataset also supports learning on these types of non-functional requirements (7).

Bag of Words and TF-IDF are used separately to answer the first research question and find the relevant features. The reason for using these methods for finding relevant features is that they work well for classifying the non-functional requirements (7, 8). Moreover, both methods cover different weighing features, and experimenting with both methods helps find more relevant features. BoW ignores the sequence of words; however, in the case of non-functional requirement classification, some algorithms can work well irrespective of the order of the information. BoW method weighs the features according to their number of occurrences in all documents (29) while the TF-IDF method gives more weight to those features which have less overall frequency but more frequency in a specific document (4, 7). Some highly relevant features are also added to help rule-based learning, part of the hybrid KNN-rule based ML algorithm.

To answer the second research question and find the ML algorithm that helps classify non-functional requirements with better accuracy, LR, SVM, MNB, KNN, the ensemble made from these algorithms, RF and hybrid KNN-rule based ML algorithms are applied. The reason for choosing these ML algorithms for classifying the non-functional requirements is because these supervised ML algorithms perform better than unsupervised or semi-supervised ML algorithms (8). The algorithms are also selected while keeping in front the dataset, which is used to classify non-functional requirements. To evaluate the accuracy of these classifiers, 10-fold cross-validation is performed. The reason for choosing this method of evaluation is that it helps in producing more accurate and unbiased results (4, 23). Furthermore, many studies adopt 10-fold cross-validation method which helps in comparing the results without bias (4, 5, 7, 23). The steps to the classification of non-functional requirements are given in **Figure 1**.

Internet of Things plays a crucial role in the health care systems of the modern era. It provides a facility to monitor, control, and prevent diseases by collecting and processing health-related data through sensors (42). Since these health care systems are critical, it is important to ensure that they work efficiently, encompassing all the required non-functional requirements. Health care systems have compromised performance or security, resulting in the loss of critical health data and even threats and attacks by hackers, which may affect the life of patients (17). An IoT-oriented health care system generally has three layers, which are the perception layer, fog layer, and cloud layer (43–45). The layered architecture is shown in **Figure 2**. The perception layer includes actuators or devices that help collect sensory data. The fog layer processes the data to produce the required response quickly. The Cloud layer deals with storing the data and big data analytics (44).

3.1. Dataset Description

In this research, the PROMISE_exp dataset (46) is used for the classification of non-functional requirements. This dataset is the expanded form of the original PROMISE dataset, which comprised only a total of 625 labeled requirement sentences (21). The expanded form contains more labeled requirement sentences including 444 functional requirements and 525 non-functional requirements (7). The 11 types of non-functional requirements are distributed among the sentences in an unbalanced fashion. **Table 2** shows the formulation of this dataset.

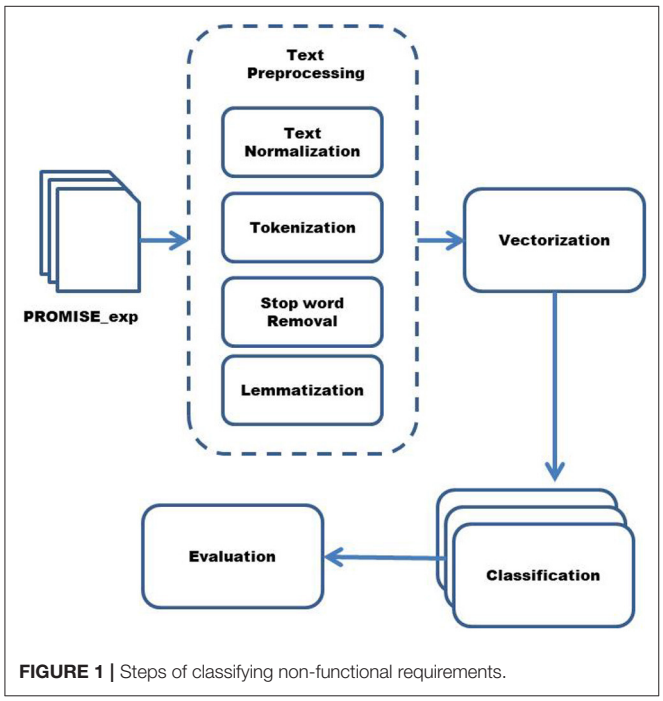


FIGURE 1 | Steps of classifying non-functional requirements.

TABLE 2 | Number of requirements per class.

Requirement type	Number of requirements
Functional requirement (FR)	444
Availability (A)	31
Security (SE)	125
Usability (US)	85
Look and Feel (LF)	49
Legal and licensing (L)	15
Maintainability (MN)	24
Operability (O)	77
Performance (PE)	67
Scalability (SC)	22
Fault Tolerance (FT)	18
Portability (PO)	12
Total	969

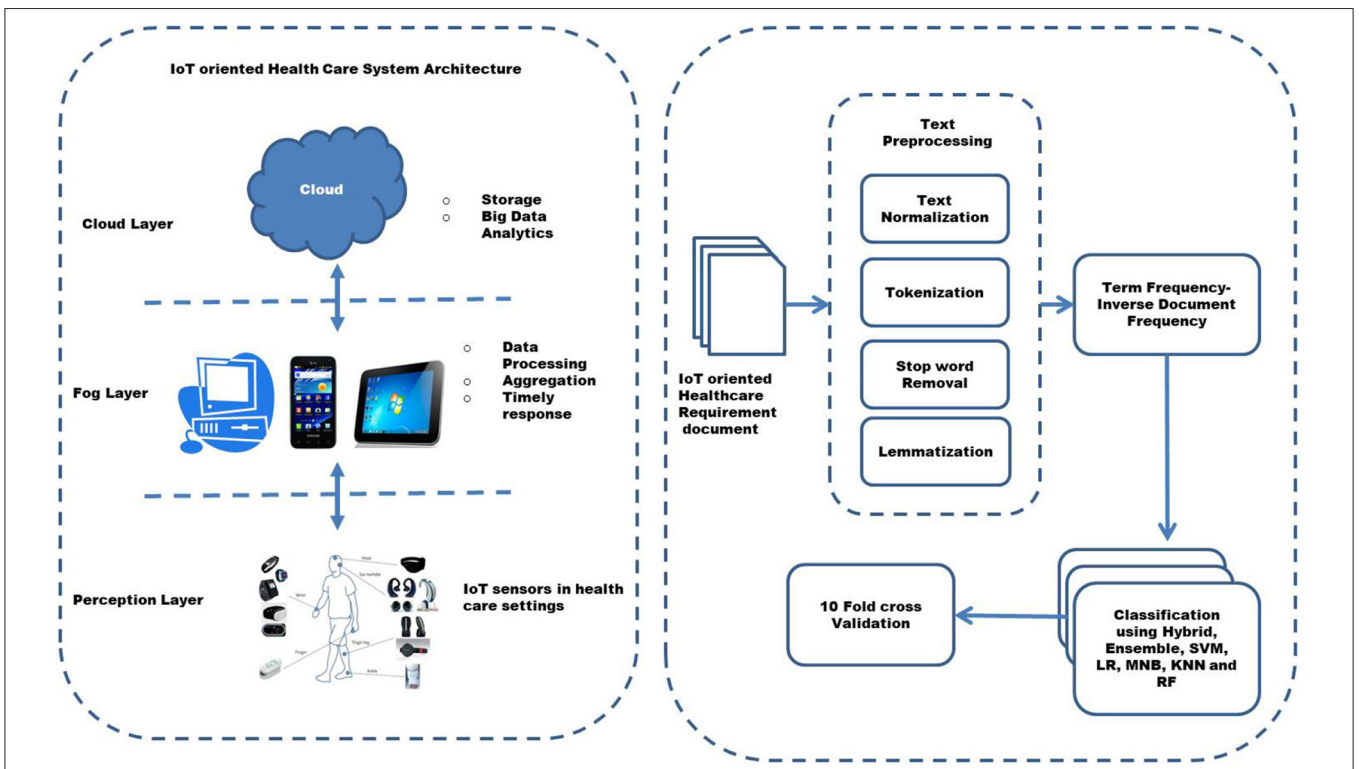


FIGURE 2 | The architectural design of Internet of Things (IoT) oriented Healthcare non-functional requirement classification.

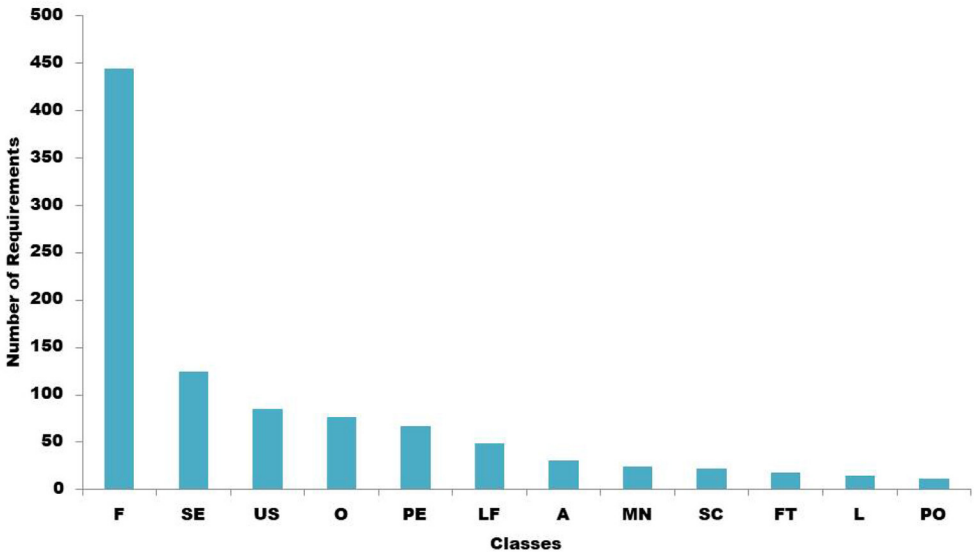


FIGURE 3 | Distribution of requirement classes.

The labeled requirements are distributed in the following manner: availability: 31, security: 125, usability: 85, look and feel: 49, legal and licensing: 15, maintainability: 24, operability: 77, performance: 67, scalability: 22, fault tolerance: 18, portability: 12, and functional requirements which are 444 in number. **Figure 3** below shows the distribution of the labeled requirement sentences.

3.2. Text Preprocessing

Text preprocessing is the first step in which the data is cleaned to remove redundant and less relevant features (8). In-text preprocessing, first, the data is normalized. During normalization, natural language text is first converted to lower case, then non-alphabetical characters, symbols, and punctuation are removed. Tokenization is performed after normalization, in which text is converted to tokens. Then in the next step, the stop words are removed, which are also less relevant features. Finally, words are lemmatized in which they are converted to their root form in order to remove extra features (7).

3.3. Feature Extraction

Feature extraction is the second step in which the preprocessed data is converted into vectors. To extract the features, BoW and TF-IDF are applied. The description of both techniques is given below.

3.3.1. Bag of Words

Bag of Words is a simple technique in which the words are converted into a numerical format based on their number of occurrences. Each feature gets its value equal to the number of times it appears in the requirement sentence. The vector of requirement sentence *'j'* is represented in Equation 1.

$$Xj = (x_{1,j} \ldots x_{i,j} \ldots xn, j) \tag{1}$$

In Equation 1, $x_{i,j}$ represents the weight of the feature which is calculated on the basis of the occurrence of *'i'* in the requirement *'j'*, whereas *'n'* represents the total number of words (7). When the vectors are created, they are given as an input to the ML algorithms in the next step.

3.3.2. Term Frequency-Inverse Document Frequency

In this technique of vectorization, two metrics play their role. The first metric is the Term Frequency (TF) which represents the number of times a particular word occurs in a requirement. The second metric is the Inverse Document Frequency which is achieved by dividing a total number of requirement sentences by the requirement occurrence number for each word and then applying a logarithmic function on the output (47). Below is the mathematical representation of TF-IDF.

$$TF - IDF(term_{i,j}) = tf_{i,j} \times idf_i \tag{2}$$

In Equation 2, $tf_{i,j}$ shows the frequency of the term i in the requirement j, where idf_i is the Inverse Document Frequency of *'i'* which is mathematically represented as Equation 3:

$$Idf_i = log \frac{total_requirements}{total_requirements_with_term_i} \tag{3}$$

3.4. Machine Learning Algorithms

The machine learning algorithms LR, SVM, MNB, KNN ensemble, RF, and hybrid KNN-rule based ML algorithms are applied to vectorized data. The classifiers are constructed not only by using the monolithic concept but also ensemble and hybrid are also considered since they yield better results according to literature (8).

3.5. Evaluation

The performance of the classifiers in terms of accuracy is evaluated using 10-fold cross-validation (7). In 10-fold cross-validation concept, the training set is divided into ten subsets of data which are of almost the exact sizes, and then testing takes place in 10 iterations, wherein each iteration, one fold which comprises 10% of the dataset is left for testing and other nine folds which comprises 90% of the dataset are used for training. In this way, each sample of the data is used once in both training and testing (30). The purpose of performing 10-fold cross-validation is to reduce the chance of biasness in results. In this research, the metric 'Accuracy' is considered to get an overall idea of the correct classification made by the classifier. Accuracy of the classifier is the measure of how many correct predictions are made by the classifier in contrast to a total number of predictions as shown in Equation 4.

$$Accuracy = \frac{(TP + TN)}{P + N} \tag{4}$$

Where TP represents the predictions correctly identified as positive, TN represents those correctly identified as negative.

4. EXPERIMENT

In order to implement the technique and perform the experiment, the Anaconda tool is used. PROMISE_exp dataset is obtained and converted into CSV format to process by the tool. The data is first preprocessed in order to remove useless features. In the first step of data preprocessing, the text is normalized. In the next step, the sentences are converted into tokens. Then in the next step, the stop words are removed. Then in the final step of text preprocessing, words are lemmatized and converted to their root form to remove extra features. After preprocessing, the data is vectorized using BoW and TF-IDF.

For classification of non-functional requirements, 7 ML algorithms are considered, which are LR, SVM, MNB, KNN, ensemble, RF, and hybrid KNN-rule based ML algorithms. The Hyperparameter of the classifiers is tested and set manually by checking their performance. The performance is evaluated in terms of accuracy by using 10-fold cross-validation.

The work in this research differs from the recent study (7) in the sense that this study considers more ML algorithms like RF, Ensemble, and hybrid KNN-rule based ML algorithm. Moreover, the relevant features are also added to increase the performance in terms of accuracy.

The non-functional requirements of an IoT oriented health care system generally include security (17, 42, 44) privacy interoperability, scalability, reliability (17, 44), accuracy, usability (17), performance (17, 44, 48), and maintainability (48). The system must possess these non-functional requirements since missing any critical non-functional requirement results in severe loss of healthcare data and other threats to patients' lives. Thus it is important to extract and classify all the non-functional requirements from the requirement document to be implemented accordingly in the IoT-oriented healthcare system. The manual process of extracting the non-functional requirements is erroneous and laborious. To avoid missing non-functional requirements, this research aims to adopt a machine learning-based approach to automatically classify the non-functional requirements from the IoT-oriented healthcare system's requirement document.

To implement the automatic technique, experimentation is performed. First dataset is created, comprising of functional and non-functional requirements of IoT-oriented health care systems (24, 25). This dataset contains requirements belonging to 8 classes: Accuracy, Reliability, Security and Privacy, Performance, Compatibility, Usability, Functional, and Maintainability. The dataset obtained is then preprocessed. IoT-oriented healthcare systems' requirement sentences are first normalized in the preprocessing stage. The text is converted to lower case in normalization, and punctuation marks and non-alphabetical characters are removed. In the second phase of preprocessing, the text is converted into tokens. In the third phase of preprocessing, stop words are removed. In the fourth phase, lemmatization converts the words into their lemma or dictionary format. After text preprocessing, vectorization is performed using TF-IDF. The features extracted by TF-IDF are then used to create classifiers by applying a machine learning algorithm. The machine learning algorithms adopted in this experiment are SVM, LR, KNN, MNB, RF, Ensemble, and hybrid KNN-rule based machine learning algorithm. For evaluation of classifiers performance, 10-fold cross-validation is performed.

5. RESULTS AND DISCUSSION

This section presents the experimentation results, which are conducted to find out the relevant features that increase the accuracy of classification of non-functional requirements (RQ1) and explore the ML algorithm, which performs better in terms of accuracy for classification of non-functional requirements (RQ2).

5.1. Relevant Features for Non-Functional Requirements Classification

In the experiment stage, two feature extraction techniques are adopted: BoW and TF-IDF. After preprocessing of text and extracting the features with the help of vectorization techniques, the top relevant features which are obtained from both the techniques are shown in **Table 3**. Furthermore, some coding is also performed to add more features.

In addition to these features obtained by the vectorization techniques, the features which are added according to the relevancy of the type of non-functional requirement are shown in **Table 4**. These added features helped achieve an accuracy as high as 85% on average, as can be seen, which is achieved by the hybrid KNN rule-based ML algorithm.

The performance of the ML algorithms when fed with the features is shown in **Figure 4**. **Figure 4** highlights the effect of BoW and TF-IDF on the classification accuracy of ML algorithms. In the hybrid KNN rule-based ML algorithm application, TF-IDF outperforms BoW to extract more relevant features and provide more accurate results. In the case of the ensemble, BoW provides more relevant features. BoW also outperforms TF-IDF in extracting more relevant features when

applied LR and MNB. In the case of SVM and KNN, the features extracted by TF-IDF proved to be more relevant. Both BoW and TF-IDF provide similar accuracy with RF.

5.2. Machine Learning Algorithm With Higher Accuracy

The second research question (RQ2) is about finding the ML algorithm that provides more accurate results than others classifying non-functional requirements. In **Figure 5**, the accuracy provided by all the 7 ML algorithms is illustrated in trend line fashion when BoW is used. There are a total of 10 iterations along which the accuracy is highlighted. The accuracy provided by the hybrid KNN rule-based ML algorithm is the highest overall. The accuracy of the ensemble drops at iteration 3, but overall, it provides almost average classification accuracy compared to the hybrid classifier. On the other hand, LR provides better accuracy than the ensemble overall but is still lower than the hybrid classifier's accuracy. SVM does not show

TABLE 3 | Top relevant features.

Top relevant features		
Number	**TF-IDF**	**BoW**
1	Must	Must
2	Information	Information
3	Sensitive	User
4	Website	Website
5	Registered	Security
6	Security	Registered
7	Secure	Secure
8	User	Sensitive
9	Third	High
10	Issue	Agreement

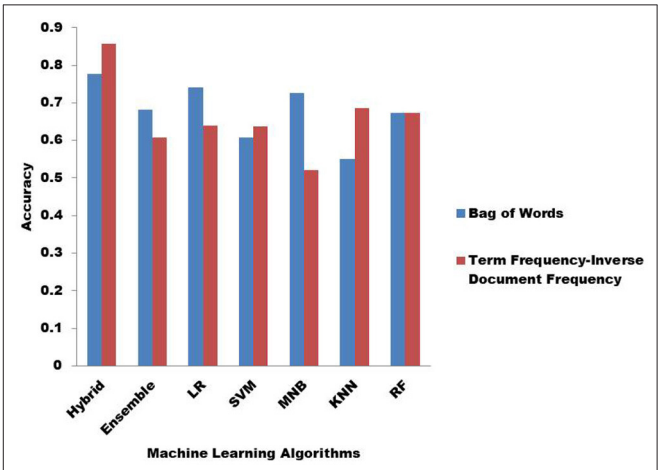


FIGURE 4 | Classification accuracy with Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF).

TABLE 4 | Additional relevant features.

Requirement type	Relevant features
Availability	availability, available, uptime, mtbf, uptime, system, support, time, hours, whenever, demand.
Legal and Licensing	licensing, legal, law, rule, rules, regulation, law, legal, claim, norm, sarbanesoxley, eula, comply, confirm, meet, violate, violating, violated, follow, agree, requirement, term, condition, accounting, standard.
Fault Tolerance	malfunction, fault, fail, failure, loss, error, warn, alive, handle, prevent, recover, retain, remain, operational, restored, system, back, connection, unavailable, operate, operating.
Security	trusted, unauthorized, malicious, hacker, illegal, secure, security, securely, limited, control, private, password, constrained, constraint, access, accessible, only, need, allow, allowed, require, required, IP address, log, logged, encrypt, encryption, hide, hidden, invisible, password, authentic, PW, attempt, unique, valid, tracking, manner, way, SSL, impose, steal, reveal, infection, virus, threat, trojan, bomb, attack.
Scalability	scale, add, accommodate, include need, increase, release, market, country, country, state, capable, supporting, new, current, customer, user, year, processing, number, grow, multiple, database server, expected, remote.
Look and Feel	look, size, color, shape, texture, position, positioned, located, location, pixel, circle, circular, square, clue, visual, scroll, tool, bar, button, icon, image, buttons, icons, menu, comply, meet, confirm, guideline, standard, display, show, illustrate, simulate, feel, sound, announce, view, grid, graphics, animation, tabbed, schema, attractive, offensive, ethnic, nomenclature, GUI, drilldown, interface.
Usability	show, display, provide, find, locate, icons, tool, tools, toolbar, image, able allow, use, user, customer, scroll, map, web, navigate, navigation, system, friendly, convention, help, ask, section, offer, customize, customization, quickly, easily, quick, easy, understand, entirely, click, select, learn, translate, language, languages, intuitive, feeling, happy, satisfied, content, free, access, accessibility, handicap, disability, one, click, familiar, self-supporting, explanatory.
Performance	perform, check, handle, every, within, second, process, processing balancing, large, simultaneously, parallel, load, response, throughput, lead, cycle, respond, minimum, fast, time, speed, rate, increase, more, high, performance, access, minute, seconds, minutes, space, memory, resource, utilize, utilization, depend, upon, later, returned, asynchronous, asynchronously, overhead, synchronize.
Maintenance	update, maintain, modify, add, integrate, remove, delete, change, replace, upgrade, easily, easy, new, different, release, updated, operate, maintenance, highly, maintainable, configurable, manage, management, function, functionality, component, module, record, data, information.
Operability	operate, exchange, interface, API, APIs, with, without, through, compatible, window, system, os, manage, format, information, data, utilize, run, current, currently, existing, equipment, environment, use, protocol, network, operated, condition, database, store, programming, language, connectivity, evoked, command, platform, prompt, service, stream, navigation, operable.
Portability	portable, portability, support, run, mobile, laptop, tablet, platform, window, os.

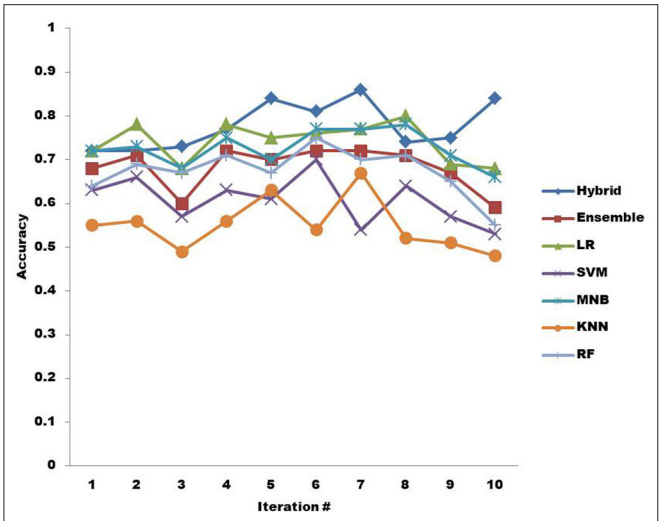


FIGURE 5 | Accuracy of machine learning algorithms with BoW.

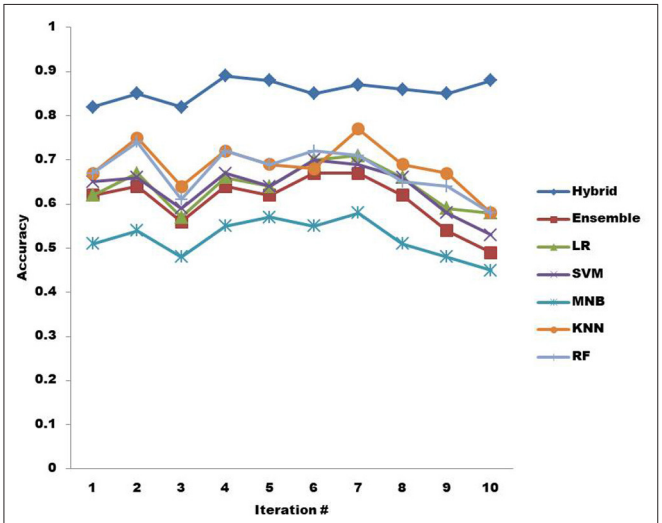


FIGURE 6 | Accuracy of machine learning algorithms with TF-IDF.

good classification accuracy and performs low overall compared to other ML algorithms except KNN. The classification accuracy provided by MNB is low compared to hybrid and LR overall but still better than other ML algorithms. KNN provides the least good accuracy compared to other ML algorithms overall. RF performs better in terms of accuracy overall than SVM and KNN.

In **Figure 6**, the classification accuracy of all ML algorithms is shown when TF-IDF is adopted. The hybrid KNN rule-based ML algorithm outperforms all other ML algorithms in providing better classification accuracy. Compared to other ML algorithms, ensemble only proved to perform better than MNB in providing better accuracy overall. LR provides better accuracy than MNB, SVM, and ensemble overall, while SVM provides better classification accuracy than just ensemble and MNB. The classification accuracy provided by MNB is low overall compared to all other ML algorithms. KNN, in this case, provides better accuracy than other ML algorithms except for hybrid. RF performs lower in accuracy than hybrid and KNN but provides better classification accuracy overall than other ML algorithms.

The classification results in terms of accuracy are shown in **Tables 5**, **6**. The result in **Table 5** clearly shows that hybrid KNN-rule based ML algorithm outperforms other ML algorithms by achieving an average accuracy of 0.778 with BoW. As a result of its nature, a hybrid classifier considers the usage of rules constructed with relevant features, which helps in the classification of non-functional requirements more accurately. Similarly, **Table 6** also shows that the highest classification accuracy among all the ML algorithms is achieved by the hybrid KNN-rule based ML algorithm, which is 0.857 accuracy on average. The results show that by using TF-IDF for feature extraction and hybrid KNN-rule based ML algorithm for classification, an average accuracy of 85.7% can be achieved, which is quite promising.

Table 5 shows the average classification accuracy of ML algorithms in 10 iterations when BoW is applied. The hybrid KNN-rule based ML algorithm achieves the highest average classification accuracy, which is 77.8%.

The results of experimentation using IoT-oriented healthcare requirement documents are shown in **Table 7**. **Table 6** shows the average classification accuracy of ML algorithms in 10 iterations when TF-IDF is applied. The hybrid KNN-rule based ML algorithm achieves the highest average classification accuracy, 85.7%.

The results show that the hybrid KNN-rule based machine learning algorithm outperforms others by showing average classification accuracy of 75.9%. In existing studies, many machine learning-based techniques are adopted to classify the non-functional requirements from the requirement document. However, to our knowledge, this is the first research that solely focuses on IoT-oriented healthcare system requirement documents to classify the non-functional requirements automatically. The automatic technique proposed in this research classifies the non-functional requirements with reasonable accuracy, enabling the development of an excellent IoT-oriented healthcare system.

Certain limitations are present in the approach presented by this research. First, the dataset adopted is minimal and contains only 104 requirements, which may affect the generalizability of the results of this research. Second, the non-functional requirements related to IoT-oriented healthcare systems covered by this research are only 7 in number. Third, the machine learning algorithms considered by this research for classification of non-functional requirements from IoT-oriented healthcare system requirement document are only 7, including only one ensemble and hybrid algorithm and five supervised machine learning algorithms. This research does not cover neural Networks, Semi-supervised or unsupervised machine algorithms.

TABLE 5 | Classification accuracy of ML algorithms using BoW.

	Classification accuracy in 10 iterations using BoW										
Machine learning algorithm	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average classification accuracy**
Hybrid	0.72	0.72	0.73	0.77	0.84	0.81	0.86	0.74	0.75	0.84	0.778
Ensemble	0.68	0.71	0.60	0.72	0.70	0.72	0.72	0.71	0.67	0.59	0.682
LR	0.72	0.78	0.68	0.78	0.75	0.76	0.77	0.80	0.69	0.68	0.741
SVM	0.63	0.66	0.57	0.63	0.61	0.70	0.54	0.64	0.57	0.53	0.608
MNB	0.72	0.73	0.68	0.75	0.70	0.77	0.77	0.78	0.71	0.66	0.727
KNN	0.55	0.56	0.49	0.56	0.63	0.54	0.67	0.52	0.51	0.48	0.551
RF	0.64	0.69	0.67	0.71	0.67	0.75	0.70	0.71	0.65	0.55	0.674

TABLE 6 | Classification accuracy of ML algorithms using Term Frequency-Inverse Document Frequency (TF-IDF).

	Classification accuracy in 10 iterations using TF-IDF										
Machine learning algorithm	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average classification accuracy**
Hybrid	0.82	0.85	0.82	0.89	0.88	0.85	0.87	0.86	0.85	0.88	0.857
Ensemble	0.62	0.64	0.56	0.64	0.62	0.67	0.67	0.62	0.54	0.49	0.607
LR	0.62	0.67	0.57	0.66	0.64	0.70	0.71	0.66	0.59	0.58	0.640
SVM	0.65	0.66	0.59	0.67	0.64	0.70	0.69	0.66	0.58	0.53	0.637
MNB	0.51	0.54	0.48	0.55	0.57	0.55	0.58	0.51	0.48	0.45	0.522
KNN	0.67	0.75	0.64	0.72	0.69	0.68	0.77	0.69	0.67	0.58	0.686
RF	0.67	0.74	0.61	0.72	0.69	0.72	0.71	0.65	0.64	0.58	0.673

TABLE 7 | Classification accuracy of ML algorithms for IoT oriented healthcare requirements.

	Classification accuracy in 10 iterations using TF-IDF										
Machine learning algorithm	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average classification accuracy**
Hybrid	0.73	0.73	0.73	1.00	0.70	0.90	0.60	0.70	0.80	0.70	0.759
Ensemble	0.82	0.64	0.27	0.64	0.80	0.60	0.70	0.70	0.60	0.50	0.627
LR	0.82	0.64	0.27	0.64	0.80	0.60	0.70	0.70	0.60	0.50	0.627
SVM	0.82	0.64	0.27	0.64	0.80	0.60	0.70	0.70	0.60	0.50	0.627
MNB	0.82	0.64	0.27	0.64	0.80	0.60	0.70	0.70	0.60	0.50	0.627
KNN	0.82	0.55	0.45	0.64	0.80	0.70	0.70	0.70	0.50	0.60	0.646
RF	0.82	0.64	0.27	0.64	0.80	0.60	0.70	0.70	0.60	0.50	0.627

6. THREATS TO VALIDITY

This section presents the possible threats which affect the validity of this research. The steps taken to mitigate the effect of threats are partially highlighted.

6.1. Construct Validity

Construct validity refers to the concept that defines the degree to which the variables measures accurately what they are supposed to measure (41). Many studies consider this measure for measuring the performance of classification, and it is believed to capture the aspect of performance required. In this research, the standard measure "Accuracy" is adopted to measure classifiers' performance.

6.2. Internal Validity

Internal validity is related to measuring the extent to which the experimental results are derived from the data and not any unconsidered variables (29). In this research, a threat is considered related to the over-fitting of test data on the machine learning algorithms. The effect of this threat is mitigated by using 10-fold cross-validation.

6.3. External Validity

External validity refers to the extent to which the results of this research apply to other settings (29). In this research, the domain and size of the dataset affect the external validity of this research. This threat is mitigated partially by selecting the PROMISE-exp dataset (46),

which contains requirement sentences from the software domain, and contains more requirements than the existing PROMISE dataset.

7. CONCLUSION AND FUTURE DIRECTIONS

Non-functional requirements are essential in the RE process since they play an essential role in driving the software architecture and determining the product quality. Since these requirements are written in natural language and often get twined up with the functional requirements, they are often missed or ignored. In order to overcome the difficulties caused due to human manual processes and inadequate tools, this research has focused on using ML algorithms for the automatic classification of non-functional requirements. Since this task of non-functional requirements classification has to be performed accurately, this study has focused on finding relevant features and ML algorithms that can perform this task with great accuracy. In this research, BoW and TF-IDF are used for feature extraction. Further, some features are also added for increasing the accuracy. The ML algorithms adopted in this research are LR, SVM, MNB, KNN, ensemble, RF, and hybrid KNN-rule based ML algorithms. The types of non-functional requirements considered are 11. The dataset used for classification is the PROMISE_exp dataset. In order to compare the accuracy of classifiers created by different ML algorithms and features, experimentation is considered. The experiment results show that by using TF-IDF for extracting the features and hybrid KNN-rule based ML algorithm for classification, an average accuracy of 85.7% can be achieved, which is a pretty excellent performance. It is noteworthy that the accuracy level achieved by this study is higher than the accuracy achieved by a recent study (7) on the PROMISE_exp dataset. Automatic classification of non-functional requirements using high-performing ML algorithms and relevant features helps the Requirement Engineers accurately classify the non-functional requirements. It saves their critical time by using fewer advanced tools and adopting human manual processes.

The research for using the machine learning approach for automatic classification of non-functional requirements with better performance is still in progress. The literature noted that the studies barely consider hybrids and ensembles to classify non-functional requirements. In the future, further study can be done to increase the size of the dataset and incorporate more types of non-functional requirements. Further attention can also be given to creating more custom classifiers by using different combinations of ML algorithms.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

IK and SI: conceptualization. SI: data curation. AJ: formal analysis, software, and methodology. IK: funding acquisition. AJ and ZJ: investigation. ZK, AA, and IK: project administration. AA, WB, and IK: resources. ZK and AJ supervision. SI, AA, and WB: validation. WB and ZK: visualization. ZJ and IK: writing—review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

REFERENCES

1. Iqbal T, Elahidoost P, Lucio L. A bird's eye view on requirements engineering and machine learning. In: *2018 25th Asia-Pacific Software Engineering Conference (APSEC)*. Nara: IEEE (2018). p. 11–20.
2. Vlas R, Robinson WN. A rule-based natural language technique for requirements discovery and classification in open-source software development projects. In: *2011 44th Hawaii International Conference on System Sciences*. Kauai, HI: IEEE (2011). p. 1–10.
3. Sabir M, Chrysoulas C, Banissi E. Multi-label classifier to deal with misclassification in non-functional requirements. In: *World Conference on Information Systems and Technologies*. Budva: Springer (2020). p. 486–93.
4. Casamayor A, Godoy D, Campo M. Identification of non-functional requirements in textual specifications: a semi-supervised learning approach. *Inf Softw Technol*. (2010) 52:436–45. doi: 10.1016/j.infsof.2009.10.010
5. Cleland-Huang J, Settimi R, Zou X, Solc P. Automated classification of non-functional requirements. *Requirements Eng*. (2007) 12:103–20. doi: 10.1007/s00766-007-0045-1
6. Winkler J, Vogelsang A. Automatic classification of requirements based on convolutional neural networks. In: *2016 IEEE 24th International Requirements Engineering Conference Workshops (REW)*. Beijing: IEEE (2016) p. 39–45.
7. Dias Canedo E, Cordeiro Mendes B. Software requirements classification using machine learning algorithms. *Entropy*. (2020) 22:1057. doi: 10.3390/e22091057
8. Binkhonain M, Zhao L. A review of machine learning algorithms for identification and classification of non-functional requirements. *Expert Syst Appl X*. (2019) 1:100001. doi: 10.1016/j.eswax.2019.100001
9. Slankas J, Williams L. Automated extraction of non-functional requirements in available documentation. In: *2013 1st International Workshop on Natural Language Analysis in Software Engineering (NaturaLiSE)*. San Francisco, CA: IEEE (2013). p. 9–16.
10. Shah A, Ahirrao S, Pandya S, Kotecha K, Rathod S. Smart cardiac framework for an early detection of cardiac arrest condition and risk. *Front Public Health*. (2021) 9:762303. doi: 10.3389/fpubh.2021.762303
11. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfik H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. *Hum Centric Comput Inf Sci*. (2020) 10:1–21. doi: 10.1186/s13673-020-00245-7

12. Awais M, Ghayvat H, Krishnan Pandarathodiyil A, Nabillah Ghani WM, Ramanathan A, Pandya S, et al. Healthcare professional in the loop (HPIL): classification of standard and oral cancer-causing anomalous regions of oral cavity using textural analysis technique in autofluorescence imaging. *Sensors*. (2020) 20:5780. doi: 10.3390/s20205780
13. Javed AR, Fahad LG, Farhan AA, Abbas S, Srivastava G, Parizi RM, et al. Automated cognitive health assessment in smart homes using machine learning. *Sustain Cities Soc*. (2021) 65:102572. doi: 10.1016/j.scs.2020.102572
14. Sarwar MU, Javed AR. Collaborative health care plan through crowdsource data using ambient application. In: *2019 22nd international multitopic conference (INMIC)*. Islamabad: IEEE (2019). p. 1–6.
15. Javed AR, Faheem R, Asim M, Baker T, Beg MO. A smartphone sensors-based personalized human activity recognition system for sustainable smart cities. *Sustain Cities Soc*. (2021) 71:102970. doi: 10.1016/j.scs.2021.102970
16. Javed AR, Sarwar MU, Khan HU, Al-Otaibi YD, Alnumay WS, et al. PP-SPA: privacy preserved smartphone-based personal assistant to improve routine life functioning of cognitive impaired individuals. *Neural Process Lett*. (2021) 1–18. doi: 10.1007/s11063-020-10414-5
17. Aruna E, Reddy A, Sunitha K. Elicitation and analysis of security requirements and patterns for IoT based health monitor. In: *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies*. Singapore: Springer (2020). p. 49–56.
18. Shabbir M, Shabbir A, Iwendi C, Javed AR, Rizwan M, Herencsar N, et al. Enhancing security of health information using modular encryption standard in mobile cloud computing. *IEEE Access*. (2021) 9:8820–34. doi: 10.1109/ACCESS.2021.3049564
19. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an IPFS-based optimization algorithm. *J Circ Syst Comput*. (2021) 31:2250010. doi: 10.1142/S0218126622500104
20. Nasir M, Javed AR, Tariq MA, Asim M, Baker T. Feature engineering and deep learning-based intrusion detection framework for securing edge IoT. *J Supercomput*. (2022) 1–15. doi: 10.1007/s11227-021-04250-0
21. Tóth L, Vidács L. Study of various classifiers for identification and classification of non-functional requirements. In: *International Conference on Computational Science and Its Applications*. Melbourne, VIC: Springer (2018). p. 492–503.
22. Hussain I, Kosseim L, Ormandjieva O. Using linguistic knowledge to classify non-functional requirements in SRS documents. In: *International Conference on Application of Natural Language to Information Systems*. London: Springer (2008). p. 287–98.
23. Rashwan A, Ormandjieva O, Witte R. Ontology-based classification of non-functional requirements in software specifications: a new corpus and svm-based classifier. In: *2013 IEEE 37th Annual Computer Software and Applications Conference*. Kyoto: IEEE (2013). p. 381–6.
24. Caballero K. *Schedule A: Functional and Non- Functional Requirements*. betterliving (2017). Available online at: http://mybetterliving.ca/wp-content/uploads/2017/01/Schedule-A-Integrated-IM-IT-Solution-for-Community-Hospice-Services-Functional-Non-Functional-Requirements-100117.pdf.
25. Sigdel S, Yadav R. *Software Requirements Specification of E-Health Architecture for Nepal* (2019) 11–4. doi: 10.13140/RG.2.2.17207.57764
26. Knauss E, Ott D. (Semi-) automatic categorization of natural language requirements. In: *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Essen: Springer (2014). p. 39–54.
27. Maiti R. Capturing. Eliciting, and prioritizing (CEP) non-functional requirements metadata during the early stages of agile software development. In: *SoutheastCon 2015*. Fort Lauderdale: IEEE (2016).
28. Yenduri G, Gadekallu TR. Firefly based maintainability prediction for enhancing quality of software. *Int J Uncertainty Fuzziness Knowl Based Syst*. (2021) 29:211–35. doi: 10.1142/S0218488521400122
29. Lu M, Liang P. Automatic classification of non-functional requirements from augmented app user reviews. In: *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*. Karlskrona (2017). p. 344–53.
30. Shreda QA, Hanani AA. Identifying non-functional requirements from unconstrained documents using natural language processing and machine learning approaches. *IEEE Access*. (2021) 1. doi: 10.1109/ACCESS.2021.3052921
31. Riaz M, King J, Slankas J, Williams L. Hidden in plain sight: automatically identifying security requirements from natural language artifacts. In: *2014 IEEE 22nd International Requirements Engineering Conference (RE)*. Karlskrona: IEEE (2014). p. 183–92.
32. Younas M, Wakil K, Jawawi D, Shah MA, Mustafa A. An automated approach for identification of non-functional requirements using word2vec model. *Int J Adv Comput Sci Appl*. (2019) 10:539–47. doi: 10.14569/IJACSA.2019.0100871
33. Kurtanovic Z, Maalej W. Automatically Classifying Functional and Non-functional Requirements Using Supervised Machine Learning. In: *2017 IEEE 25th International Requirements Engineering Conference (RE)*. Lisbon: IEEE (2017). p. 490–5.
34. Iwendi C, Jalil Z, Javed AR, Reddy T, Kaluri R, Srivastava G, et al. Keysplitwatermark: zero watermarking algorithm for software protection against cyber-attacks. *IEEE Access*. (2020) 8:72650–60. doi: 10.1109/ACCESS.2020.2988160
35. Srivastava G, Deepa N, Prabadevi B, Reddy MPK. An ensemble model for intrusion detection in the internet of softwarized things. In: *Adjunct Proceedings of the 2021 International Conference on distributed Computing and Networking*. Nara (2021). p. 25–30.
36. Pirbhulal S, Pombo N, Felizardo V, Garcia N, Sodhro A, Mukhopadhyay SC. Towards machine learning enabled security framework for iot-based healthcare. In: *2019 13th International Conference on Sensing Technology (ICST)*. Sydney, NSW: IEEE (2019). p. 1–6.
37. Bharadwaj HK, Agarwal A, Chamola V, Lakkaniga NR, Hassija V, Guizani M, et al. A review on the role of machine learning in enabling IoT based healthcare applications. *IEEE Access*. (2021) 9:38859–90. doi: 10.1109/ACCESS.2021.3059858
38. Newaz AI, Sikder AK, Rahman MA, Uluagac AS. Healthguard: a machine learning-based security framework for smart healthcare systems. In: *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. Granada: IEEE (2019). p. 389–96.
39. Goundar S. *Research Methodology and Research Method*. Wellington: Victoria University of Wellington (2013).
40. Deist T, Patti A, Wang Z, Krane D, Sorenson T, Craft D. Simulation-assisted machine learning. *Bioinformatics*. (2019) 35:4072–80. doi: 10.1093/bioinformatics/btz199
41. Mahmoud A, Williams G. Detecting, classifying, and tracing non-functional software requirements. *Requirements Eng*. (2016) 21:357–81. doi: 10.1007/s00766-016-0252-8
42. Gelogo YE, Hwang HJ, Kim HK. Internet of things (IoT) framework for u-healthcare system. *Int J Smart Home*. (2015) 9:323–30. doi: 10.14257/ijsh.2015.9.11.31
43. Mohiyuddin A, Javed AR, Chakraborty C, Rizwan M, Shabbir M, Nebhen J. Secure cloud storage for medical IoT data using adaptive neuro-fuzzy inference system. *Int J Fuzzy Syst*. (2021) 1–13. doi: 10.1007/s40815-021-01104-y
44. Calvillo-Arbizu J, Román-Martínez I, Reina-Tosina J. Internet of things in health: requirements, issues, and gaps. *Comput Methods Progr Biomed*. (2021) 208:106231. doi: 10.1016/j.cmpb.2021.106231
45. Ikram AA, Javed AR, Rizwan M, Abid R, Crichigno J, Srivastava G. Mobile cloud computing framework for securing data. In: *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*. Brno: IEEE (2021). p. 309–15.
46. Lima M, Valle V, Costa E, Lira F, Gadelha BF. Software engineering repositories: expanding the promise database. In: *Proceedings of the XXXIII Brazilian Symposium on Software Engineering*. Salvador (2019).
47. Amasaki S, Leelaprute P. The Effects of Vectorization Methods on Non-Functional Requirements Classification. In: *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. Prague (2018). p. 175–82.
48. Nasiri S, Sadoughi F, Tadayon MH, Dehnad A. Security requirements of internet of things-based healthcare system: a survey

study. *Acta Inf Med.* (2019) 27:253. doi: 10.5455/aim.2019.27.253-258

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors AJ.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Khurshid, Imtiaz, Boulila, Khan, Abbasi, Javed and Jalil. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 25 March 2022
doi: 10.3389/fpubh.2021.803659

Analysis of Factors Affecting Cranial Nerve Function of Patients With Vascular Mild Cognitive Impairment Through Functional Magnetic Resonance Imaging Under Artificial Intelligence Environment

Lifang Zhang[1,2*†], *Yanran Li*[3†], *Lin Bian*[1], *Qingrong Luo*[1], *Xiaoxi Zhang*[2] *and Bing Zhao*[1]

[1] *Department of Neurology, Changzhi People's Hospital, Changzhi Medical College, Changzhi, China,* [2] *Department of Mental Health, Changzhi Medical College, Changzhi, China,* [3] *Department of Radiology, First Affiliated Hospital of Xinjiang Medical University, Urumqi, China*

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Zhaoyang Han, University of Aizu, Japan
Abdul Rehman Javed, Air University, Pakistan
Praveen Kumar, VIT University, India

***Correspondence:**
Lifang Zhang
zlf_7296@163.com

[†] *These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *28 October 2021*
Accepted: *22 November 2021*
Published: *25 March 2022*

Citation:
Zhang L, Li Y, Bian L, Luo Q, Zhang X and Zhao B (2022) Analysis of Factors Affecting Cranial Nerve Function of Patients With Vascular Mild Cognitive Impairment Through Functional Magnetic Resonance Imaging Under Artificial Intelligence Environment. Front. Public Health 9:803659. doi: 10.3389/fpubh.2021.803659

The study aimed to explore the risk factors of effects of patients with vascular mild cognitive impairment (VaMCI) through functional magnetic resonance imaging (fMRI). In this study, 62 patients were selected from the department of neurology, admitted to Changzhi People's Hospital from October 1, 2018 to February 1, 2020. Patients with VaMCI were defined as the VaMCI group according to Clinical Dementia Rating (CDR), and subjects with normal cognitive function were defined as the normal control (NC) group. All patients underwent fMRI to identify the amplitude low-frequency fluctuation (ALFF) and regional homogeneity (ReHo) values, and to analyze their association with VaMCI. The results showed that the VaMCI group had lower scores for Mini-mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and their subitems (visual space and execution, recall, attention and computation, and language ability) than NC group, with statistical differences ($P < 0.05$). In VaMCI group, the brain regions with increased ALFF values were the left temporal lobe, left parietal lobe, right temporal lobe, right parietal lobe, and posterior cingulate gyrus. Of them, the left parietal lobe and right temporal lobe were negatively correlated with the recall score on MMSE scale ($r = -0.216$, $r = -0.132$, $P < 0.01$). In VaMCI group, the brain regions with decreased ReHo values were the left temporal lobe, occipital lobe, and left middle temporal gyrus. Of them, the left temporal lobe and occipital lobe were positively correlated with MoCA score ($r = 0.473$, $r = 0.848$, $P < 0.01$). In conclusion, VaMCI patients have cognitive impairment and abnormally increased spontaneous brain activity, especially in the left parietal lobe and the right temporal lobe. At rest, VaMCI patients show decreased whole-brain ReHo in the left medial temporal lobe and occipital lobe. Hypertension is a high-risk factor for cognitive impairment in VaMCI patients. The study can provide a theoretical basis for early diagnosis of VaMCI.

Keywords: artificial intelligence, vascular mild cognitive impairment, computer functional magnetic resonance imaging, neuropsychological assessment, whole brain regional homogeneity

INTRODUCTION

Cerebrovascular diseases are a main factor causing vascular cognitive impairment (VCI) (1, 2). With the population aging, the incidence of VCI increases significantly, second only to Alzheimer's disease (3–5). Similar to other cognitive diseases, VCI is characterized by gait, emotion, behavior, and urination disorders (6, 7). Cognitive impairment in an early stage is often overlooked, while it is difficult to treat once it develops to the late stage, such as dementia (8). Vascular mild cognitive impairment (VaMCI), as a prodromal stage of VCI, is closely related to executive dysfunction (9, 10), so early detection of VaMCI is particularly important.

Currently, the clinical diagnosis of VaMCI mainly relies on neurological function evaluation and imaging examination. Of many diagnostic methods, artificial intelligence-based imaging is of great significance for early diagnosis of VaMCI (11–14). With the continuous improvement of modern medical technology, artificial intelligence (AI) technology is gradually introduced into fMRI and plays an important role. AI can mine effective imaging information from big data, providing convenience for doctors in the diagnosis and treatment of cerebrovascular diseases (15–18). The fMRI imaging technology is now widely used to explore the mechanism of cerebrovascular cognitive function by detecting the activity of brain neurons. When different activities occur, both vascular cognitive function and neuronal activity have corresponding changes (19). Artificial intelligence-based fMRI technology features non-invasiveness, repeatable and high spatial resolution. It is an important means to study the neural mechanism of VCI patients (20–23). However, there are few studies to use fMRI technology to diagnose VaMCI, and the related research is still in the exploratory stage at present.

In this study, case data and fMRI data of patients in the department of neurology and the department of physical examination, admitted to Changzhi People's Hospital from October 1, 2018 to February 1, 2020, were collected, to identify the ALFF and ReHo value, and to analyze their association with VaMCI, expected to provide a basis for the early diagnosis of cerebrovascular diseases.

MATERIALS AND METHODOLOGY

In this study, the subjects underwent fMRI to identify the ALFF and ReHo values, and to analyze their association with VaMCI, aiming to explore the risk factors of VaMCI by fMRI combined with neuropsychology under artificial intelligence.

Basic Information of Research Subjects

Case data and fMRI data of patients in the department of neurology and the department of physical examination, admitted to Changzhi People's Hospital from October 1, 2018 to February 1, 2020, were collected. Thirty two patients were in the VaMCI group according to the diagnostic criteria of CDR and the Guidelines for the Diagnosis and Treatment of Vascular Cognitive Impairment in China in 2019 (24). A total of 30 healthy patients without memory loss were in the NC group. The subjects in the NC group had no other related disorders that might affect cognitive function.

Inclusion criteria: 1. patients with complete case history and fMRI data; 2. acute cerebrovascular diseases did not occur within at least 1 year; 3. patients diagnosed with VCI as per the criteria of *the Guidelines for the Diagnosis and Treatment of Vascular Cognitive Disorders in China* in 2019; 4. there were cerebrovascular risk factors, and no evidence of cerebrovascular diseases such as hydrocephalus, intracerebral hemorrhage, lacunar cerebral infarction, and leukodystrophy from fMRI; and 5. other intracranial lesions were excluded, such as: tumor, trauma, infection or surgery.

Exclusion criteria: 1. Intracranial vascular color doppler ultrasound, head fMRI or head CT angiography showed intracranial vascular stenosis of more than 50%; 2. patients with abnormal neurological diseases such as Parkinson's disease, brain tumor, and encephalitis; 3. patients unable to complete the examination due to moderate aphasia, depression, or mental disorder; 4. patients with genetic microvascular diseases; and 5. patients with a history of major depression, drug, alcohol, or drug dependence.

Neurological Function Evaluation

The environment was quiet where the patient was tested. The test scales include MMSE, MoCA, Activities of Daily Living (ADL), and CDR (25), factoring into subitems such as memory, numeral symbol conversion, attention and computational power, visual space and executive function, recall power, orientation, language ability, and abstract ability. Neurological function evaluation was performed by neurologists with years of experience.

MRI Examination

MRI scanner used the Siemens verio 3.0T computer. The patient was in a supine position with his head in his hands. The patient should breath peacefully but not to think or fall asleep. For patients unable to cooperate in the MRI examination, chloral hydrate can be taken orally for sedation. The patient was scanned for conventional horizontal position, axial T2WI, vector T1WI, coronary T2, and fluid attenuated inversion recovery (FLAIR) images, and scanning parameters are set as follows:

I. Conventional scanning: scanning voltage of 120 kV, current of 150mA, with T1WI repetition time (TR) of 1800 ms, time of echo (TE) of 25 ms, T2WI (TR/TE = 2800/75 ms), FLAIR (TR/TE = 7500/120 ms), matrix 525 × 525, scanning layer thickness of 4 mm, and layer spacing of 5 mm.
II. fMRI scanning: TR/TE = 1800/30 ms, matrix 70 × 70, scanning layer thickness of 3 mm, layer distance of 1 mm. there are a total of 250 time points in 10 min.

Pretreatment and Analysis of fMRI Image Data

The MRI data are processed by Statistic parameter Mapping (SPM8) software of Matlab platform, and the specific steps are shown below (26).

I. Data conversion: SPM8 software converts the original data Dicom into NII format.

II. Time correction: all images are obtained at the same time to eliminate the influence of time difference on the data.
III. Head motion correction: during the scan, the patient's breathing and heartbeat will have certain influence on the head, so it is necessary to keep head motion within a maximum shift of 3 mm.
IV. Space standardization: due to brain morphology differences, space standardization is required. The volume size is set as 3 ×3 ×3 mm for resampling.
V. Spatial smoothing: spatial processing with 4 × 4 × 4 mm full width at half maximum (FWHM) is performed to suppress high frequency signal and image noise.
VI. De-linear drift and low frequency filtering: in order to eliminate baseline drift, it is necessary to remove linear drift, and the frequency band of 0.01∼ 0.09 hz is used for low-frequency filtering to remove high-frequency interference.
VII. Calculation of ALFF: the ALFF value indicates the activity of voxel in the brain, which is positively correlated with spontaneous brain activity. DPARSF software is used to calculate the average amplitude value between 0.01 and 0.09 Hz.

Calculation of ReHo

A high ReHo value indicates consistent local neuron signal activity and a low value indicates decreased consistency. The ReHo value of each voxel is obtained by computer analysis, and the ReHo values of all brain regions can form a complete ReHo diagram.

Statistical Analysis

SPSS13.0 statistical software is used to analyze and process all experimental data. The count experimental data are expressed as a percentage (%) and the measurement data are expressed as mean ± standard deviation (± s). The $\chi 2$ test is used for the gender and education background analyses of the patients. *T*-test is used for age comparison between the two groups. The fMRI data are processed by DPARSF and SPM8 software, and the differences in ALFF and ReHo are analyzed by *t*-test. SPSS Pearson software is employed to analyze the correlation between ALFF, ReHo with MMSE, MoCA. $P < 0.05$ indicates statistically significant differences.

RESULTS

Basic Information

Figure 1 showed the gender and age of patients in NC group and VaMCI group. In the NC group, there were 17 males (56.7%) and 13 females (43.3%). In the VaMCI group, there were 18 males (56.3%) and 14 females (43.7%). The χ^2-test results showed no significant differences in gender ($P > 0.05$). The subjects in NC group were aged between 45–85 years old and averaged 65.3 ± 9.38; the patients in VaMCI group were aged between 50–88 years old, and averaged 69.54 ± 7.23. There was no significant difference in age after *t*-test ($P > 0.05$). The basic information was comparable in gender and age ($P > 0.05$). In terms of the education background of the two groups, 5 cases of illiteracy in NC group accounted for 16.7%, 6 cases of primary school accounted for 20%, 13 cases of junior high school accounted for 43.3%, and 6 cases of higher education accounted for 20%. In VaMCI group, 3 cases of illiteracy accounted for 9.4%, 4 cases of primary school accounted for 12.5%, 15 cases of junior high school accounted for 46.9%, and 10 cases of higher education accounted for 31.2%. The χ^2-test showed no significant difference in educational background ($P > 0.05$).

The Medical History and Family History

Figure 2 showed the medical history and family history of patients in NC group and VaMCI group. In NC group (30 cases), 10 cases of hypertension accounted for 33.3%, 8 cases of diabetes accounted for 26.7%, 8 cases of heart disease accounted for 26.7%, 9 cases of hyperlipidemia accounted for 30%, and 3 cases with a family history of dementia accounted for 10%. In VaMCI group (32 cases), there were 17 cases with hypertension (53.1%), 15 patients with diabetes (46.9%), 15 patients with heart disease (46.9%), 7 patients with hyperlipidemia (21.9%), and 0 patients with a family history of dementia. The number of hypertension

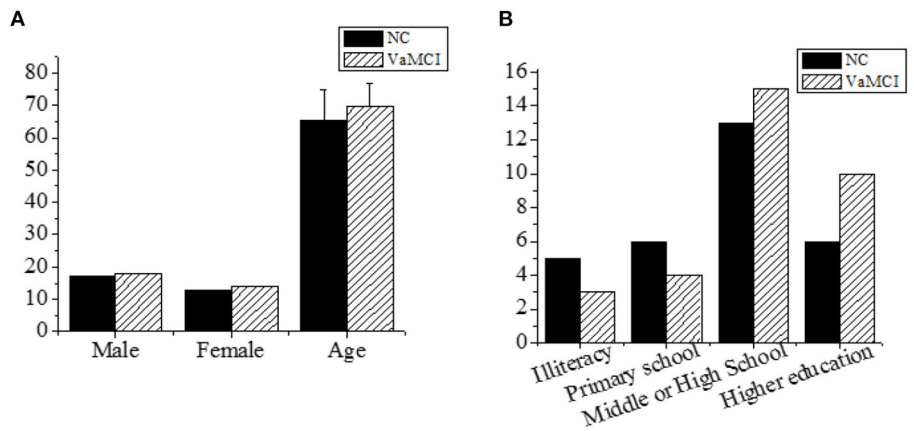


FIGURE 1 | The basic information. **(A)** The age and gender; **(B)** the education background.

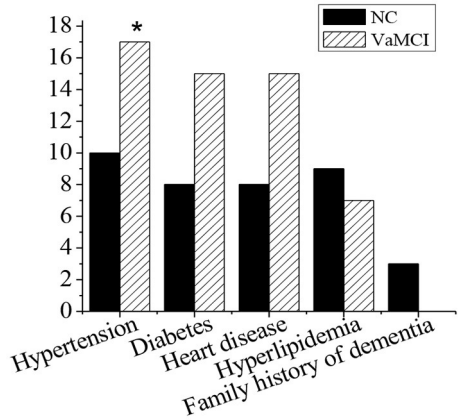


FIGURE 2 | Medical history and family history. There was a statistically significant difference in hypertension between the VaMCI group and the control group (*$P < 0.05$).

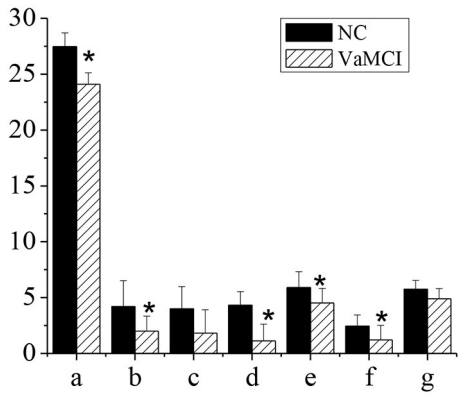


FIGURE 3 | MMSE and subitem scales. The comparison between the NC group and the VaMCI group was statistically significant, *$P < 0.05$. a-MMSE; b-visual space and execution; c-memory; d-recall power; e-attention and computation; f-language ability; g-orientation.

patients in VaMCI group was significantly higher than that in the control group (*$P < 0.05$).

MMSE and Subitem Scores

Figure 3 showed the MMSE and subitem acores of patients in NC group and VaMCI group. The total MMSE scores of VaMCI group were (24.11 ± 1.01), the subitem scores were (2.01 ± 1.34) for spatial and executive ability, (1.12 ± 1.5) for recall ability, (4.52 ± 1.3) for attention and computational ability, and (1.22 ± 1.3) for language ability, all of which were lower vs. the NC group, where the total MMSE score was (27.46 ± 1.23), and the subitem scores were (4.2 ± 2.33) for spatial and executive ability, (4.33 ± 1.21) for recall power, (5.89 ± 1.44) for attention and computational ability, and (2.45 ± 0.99) for language ability ($P < 0.05$). There was no statistically significant difference in memory and orientation between the NC group and the VaMCI group ($P > 0.05$).

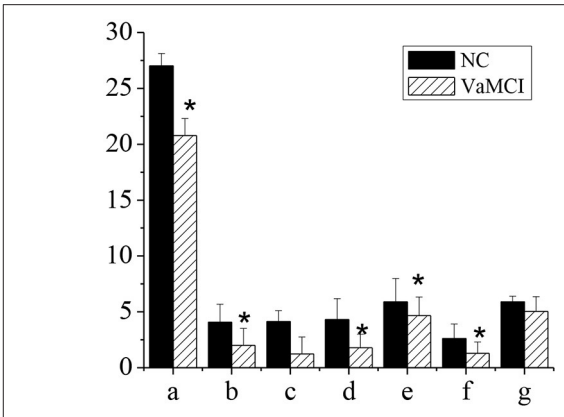


FIGURE 4 | MoCA and subitem scores. The comparison between the NC group and the VaMCI group was statistically significant *$P < 0.05$. a-MMSE; b-visual space and execution; c-memory; d-recall power; e-attention and computation; f-language ability; g-orientation.

TABLE 1 | Numerical analysis of ALFF in each brain region of VaMCI group.

Brain area	MINI coordinate			*T* value	Voxel
	x	y	z		
Left temporal lobe	−46	−15	29	−2.98	76
Left occipital lobe	58	21	42	−3.67	42
Left parietal lobe	−18	30	17	+2.54	52
Right temporal lobe	−62	−18	40	−4.28	61
Right occipital lobe	47	71	27	−3.17	30
Right parietal lobe	−8	−46	59	+3.26	51
Posterior cingulate	11	−51	25	+3.54	65

MNI, human brain region coordinates; T value, activation degree.

MoCA and Subitem Scores

Figure 4 showed the MoCA and subitem scores of patients in NC group and VaMCI group. The total MoCA score of VaMCI group was (20.78 ± 1.52), the subitem scores were (2.01 ± 1.51) for spatial and executive ability, (1.79 ± 1.22) for recall power, (4.66 ± 1.67) for attention and computational ability, and (1.29 ± 1.01) for language ability, all of which were lower vs. the NC group, where the total MoCA score was (27.01 ± 1.12), and the subitem scores were (4.08 ± 1.61) for spatial and executive ability, (4.3 ± 1.89) for recall power, (5.89 ± 2.1) for attention and computational ability, and (2.61 ± 1.3) for language ability ($P < 0.05$). There was no statistically significant difference in memory and orientation between the NC group and the VaMCI group ($P > 0.05$).

The fMRI Characteristics

Table 1 showed the ALFF value of each brain region in the VaMCI group. When the number of continuous volume pixel is greater than 50, the cluster is defined as a brain region, as shown in **Figure 5**. Compared with the NC group, there were 7 clusters with increased ALFF values in VaMCI group, which were located in the left temporal lobe, left parietal lobe, right temporal lobe,

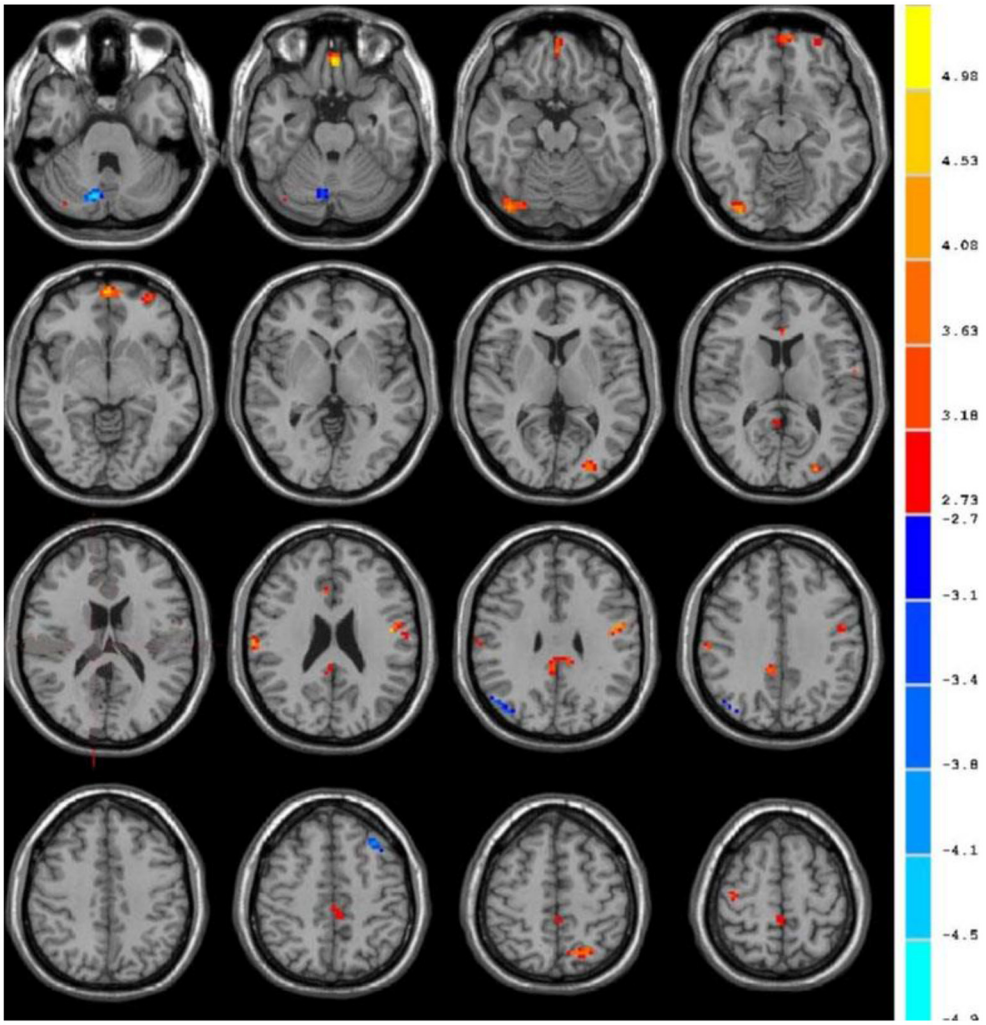


FIGURE 5 | fMRI images to show ALFF. The blue area represented the brain areas with increased ALFF in VaMCI group, the orange area represented the brain areas negatively correlated with the scale score, and the color bar represented the size of T value.

right parietal lobe, and posterior cingulate gyrus, with statistical significances ($P < 0.05$). No statistical differences were noted in brain areas of the left occipital lobe and the right occipital lobe ($P > 0.05$), and there was no brain area with abnormally decreased ALFF.

ReHo Values

Also, the cluster is defined as a brain region when the number of continuous voxels is greater than 50, as shown in **Figure 6**. **Table 2** showed the ReHo value of each brain area in VaMCI group. Compared with the NC group, there were 6 clusters with decreased ReHo values in VaMCI group, which were located in the left temporal lobe, occipital lobe, and left middle temporal gyrus, with statistical differences ($P < 0.05$). No statistical difference was noted in the left inferior temporal gyrus, left Para hippocampal gyrus, and insula ($P > 0.05$). There were 4 clusters with increased ReHo values, which were the left superior temporal gyrus, the right frontal lobe, the left frontal lobe, and the left inferior frontal gyrus ($P < 0.05$).

Spontaneous Brain Activity and Cognitive Impairment in VaMCI Group

The correlation between the ALFF value and MMSE score was analyzed. As shown in **Figure 7**, the right temporal lobe and the left parietal lobe were negatively correlated with the recall power of MMSE scale ($r = -0.216$, $r = -0.132$, $P < 0.01$), and the difference was statistically significant.

Correlation of ReHo With Cognitive Impairment

The correlation between the ReHo value and the MoCA score was analyzed. As shown in **Figure 8**, the ReHo in

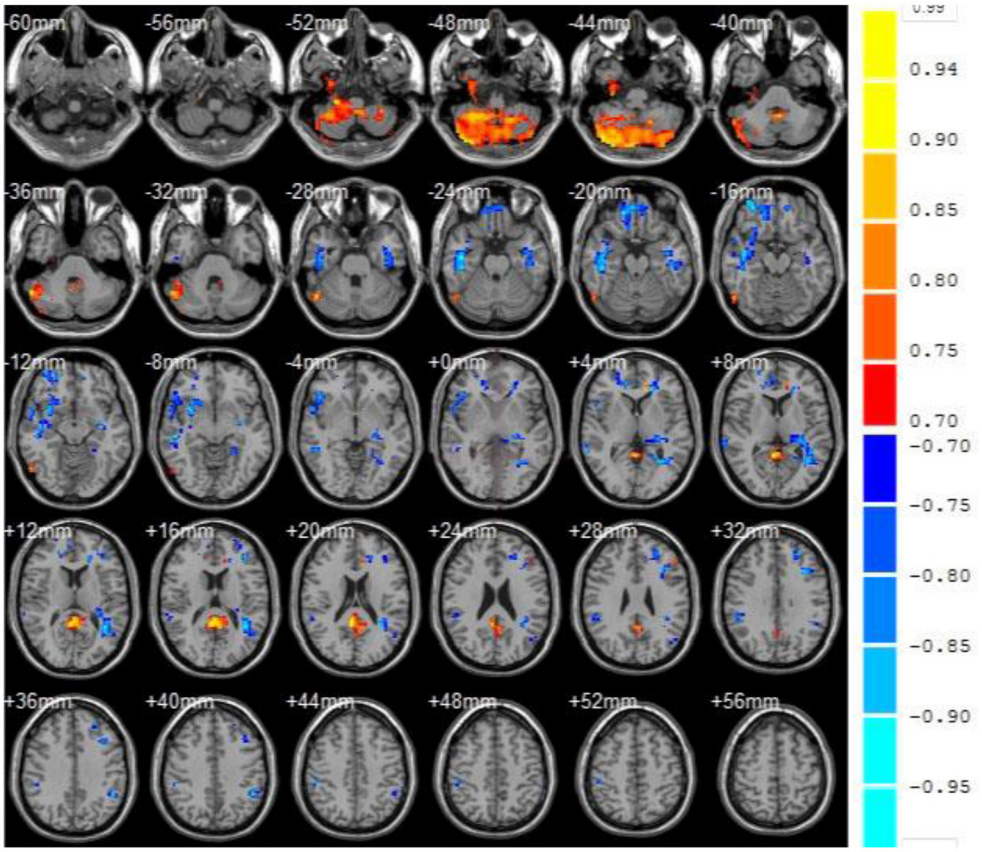


FIGURE 6 | fMRI images to show ReHo. Compared with the NC group, the blue area represented the brain areas with decreased ReHo value in VaMCI group, the orange area represented the brain areas negatively correlated with the score, and the color bar represented the size of T value.

TABLE 2 | Numerical analysis of ReHo in different brain regions of VaMCI group.

Brain region	MINI coordinate			*T* value	Voxel
	x	y	z		
Left temporal lobe	71	−59	−24	+2.98	74
Occipital lobe	−43	−64	−11	−3.67	53
Left inferior temporal gyrus	−46	−43	−10	+2.54	36
Left middle temporal gyrus	56	−35	−6	+4.28	84
Left Para hippocampal gyrus	47	47	21	−3.17	48
Insular lobe	26	8	5	+3.26	26
Left superior temporal gyrus	−44	5	−4	−2.89	58
Right frontal lobe	−14	16	43	+3.75	84
Left frontal lobe	−17	25	−7	−3.12	69
Left inferior frontal gyrus	−37	2	19	−2.91	52

MNI, human brain region coordinates; T value: activation degree.

the left temporal lobe and occipital lobe was positively correlated with the recall power of the MoCA scale ($r = 0.473$, $r = 0.848$, $P < 0.01$), and the difference was statistically significant.

DISCUSSION

VCI has a serious impact on human health and life. VCI at an early stage is often ignored thanks to mild symptoms. When it develops into dementia, it is already too late and difficult to treat. Therefore, early detection of VaMCI is particularly important (27). In this study, case data and fMRI data based on AI technology of patients in department of neurology and department of physical examination, admitted to Changzhi People's Hospital hospital from October 1, 2018 to February 1, 2020 were collected. The total MMSE score of VaMCI group was (24.11 ± 1.01), the subitem scores were (2.01 ± 1.34) for spatial and executive ability, (1.12 ± 1.5) for recall ability, (4.52 ± 1.3) for attention and computational ability, and (1.22 ± 1.3) for language ability, all of which were lower vs. the NC group, where the total MMSE score was (27.46 ± 1.23), and the subitem scores were (4.2 ± 2.33) for spatial and executive ability, (4.33 ± 1.21) for recall power, (5.89 ± 1.44) for attention and computational ability, and (2.45 ± 0.99) for language ability ($P < 0.05$). There was no statistically significant difference in memory and orientation between the NC group and the VaMCI group ($P > 0.05$). It indicated that VaMCI patients had cognitive impairment, consistent with previous research (28). The total MoCA score of

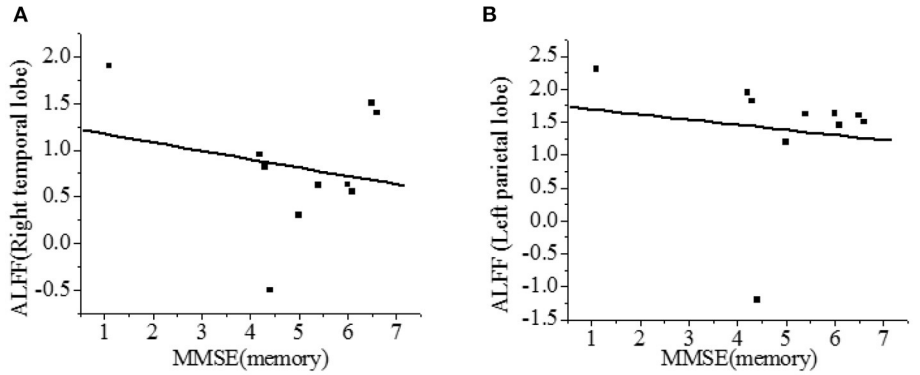


FIGURE 7 | Correlation of ALFF with recall power of MMSE scale [**(A)**-right temporal lobe; **(B)**-left parietal lobe]. There was a significant difference, $P < 0.01$.

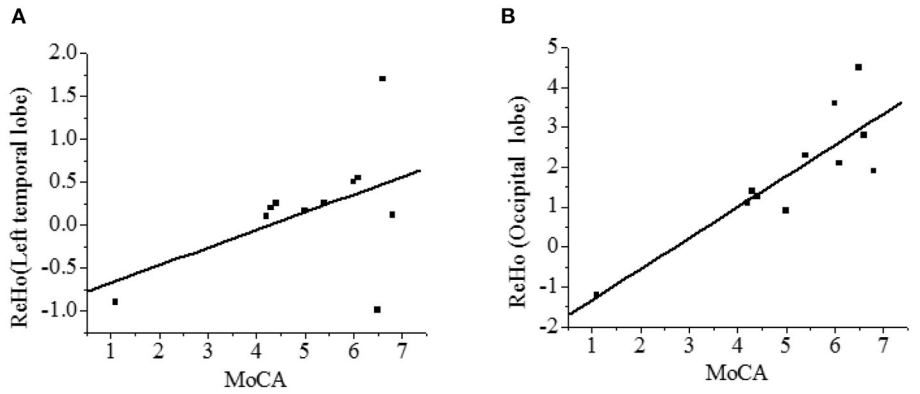


FIGURE 8 | Correlation analysis between ReHo and MoCA score [**(A)**-left temporal lobe; **(B)**-occipital lobe]. There was a statistical significance, $P < 0.01$.

VaMCI group was (20.78 ± 1.52), the subitem scores were (2.01 ± 1.51) for spatial and executive ability, (1.79 ± 1.22) for recall power, (4.66 ± 1.67) for attention and computational ability, and (1.29 ± 1.01) for language ability, all of which were lower vs. the NC group, where the total MoCA score was (27.01 ± 1.12), and the subitem scores were (4.08 ± 1.61) for spatial and executive ability, (4.3 ± 1.89) for recall power, (5.89 ± 2.1) for attention and computational ability, and (2.61 ± 1.3) for language ability ($P < 0.05$). There was no statistically significant difference in memory and orientation between the NC group and the VaMCI group ($P > 0.05$). It indicated that memory and orientation in VaMCI group were less impaired than other subitems. Studies have confirmed that the damage of subfrontal cortical neurons in the brain was the main cause of VaMCI (29).

In this study, the ALFF value of fMRI image was used to study the local spontaneous brain activity of patients in VaMCI group. Compared with the NC group, there were 7 clusters with increased ALFF values in VaMCI group, which were located in the left temporal lobe, left parietal lobe, right temporal lobe, right parietal lobe, and posterior cingulate gyrus, with statistical significances ($P < 0.05$). No statistical differences were noted in brain areas of the left occipital lobe and the right occipital lobe ($P > 0.05$), and there was no brain area with abnormally decreased ALFF. It showed that the spontaneous activity was increased in the left temporal lobe, the left parietal lobe, the right temporal lobe, the right parietal lobe, and the posterior cingulate gyrus in the rest state of patients in VaMCI group. The right temporal lobe and the left parietal lobe were negatively correlated with recall power on MMSE scale ($r = -0.216$, $r = -0.132$, $P < 0.01$). The parietal lobe played an important role in controlling visual movement, eye movement, and attention (30). The analysis of ALFF in VaMCI group showed that the temporal lobe and the Para hippocampal gyrus were closely related to memory, and that episodic memory was mainly impaired in early VaMCI period.

Compared with the NC group, there were 6 clusters with decreased ReHo values in VaMCI group, which were located in the left temporal lobe, occipital lobe, and left middle temporal gyrus, with statistical differences ($P < 0.05$). No statistical difference was noted in the left inferior temporal gyrus, left Para hippocampal gyrus, and insula ($P > 0.05$). There were 4 clusters with increased ReHo values, which were the left superior temporal gyrus, the right frontal lobe, the left frontal lobe, and the left inferior frontal gyrus ($P < 0.05$). Temporal lobe and occipital lobe are important brain areas involved in the subcortical circuits of the brain. The middle temporal gyrus is involved in cognitive and memory functions (31), and the

occipital lobe is responsible for processing of language, motion sensation, and abstract concepts (32). Chen et al. used fMRI to study the ReHo value of VaMCI patients in resting state. The findings were consistent with the conclusions in this study (33), indicating that the spontaneous brain activity was reduced in these areas mentioned above, reflecting the overall cognitive impairment of patients to a certain extent. In a word, VaMCI patients have cognitive impairment and abnormally increased spontaneous brain activity, especially in the left parietal lobe and the right temporal lobe. At rest, VaMCI patients show decreased whole-brain ReHo in the left medial temporal lobe and occipital lobe. Hypertension is a high-risk factor for cognitive impairment in VaMCI patients. The study can provide a theoretical basis for early diagnosis of VaMCI.

CONCLUSION

In the study, the fMRI was performed on subjects to identify the ALFF and Reho values and to analyze their association with VaMCI. It can be concluded that VaMCI patients have cognitive impairment and abnormally increased spontaneous brain activity, especially in the left parietal lobe and the right temporal lobe. At rest, VaMCI patients show decreased whole-brain ReHo in the left medial temporal lobe and occipital lobe. Hypertension is a high-risk factor for cognitive impairment in VaMCI patients. However, some limitations in the study should be noted. The sample size is small, which will reduce the power of the study. In the follow-up, an expanded sample size is necessary to strengthen the findings of the study. All in all, the study can provide a theoretical basis for early diagnosis of VaMCI.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Neurology, Changzhi People's Hospital, Changzhi Medical College. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

LZ wrote the manuscript. YL, LB, and QL collected and analyzed general data of patients. XZ and BZ were responsible for Follow-up indexes analysis. BZ helped with statistical analysis. All authors read and approved the final manuscript.

FUNDING

This work was supported by Scientific Research Project of Shanxi Provincial Health Commission in 2021 (No. 2021012). The key special project of "four batches" of science and technology innovation plan of Shanxi Provincial Health Commission (No. 2020XM38).

REFERENCES

1. Fischer N, Weber B, Riechelmann H. Presbyakusis - Altersschwerhörigkeit [Presbycusis - age related hearing loss]. *Laryngorhinootologie.* (2016) 95:497–510. doi: 10.1055/s-0042-106918
2. Watson N, Ding B, Zhu X, Frisina RD. Chronic inflammation - inflammaging - in the ageing cochlea: a novel target for future presbycusis therapy. *Ageing Res Rev.* (2017) 40:142–8. doi: 10.1016/j.arr.2017.10.002
3. Young YH. Contemporary review of the causes and differential diagnosis of sudden sensorineural hearing loss. *Int J Audiol.* (2020) 59:243–53. doi: 10.1080/14992027.2019.1689432
4. Tadros SF, D'Souza M, Zhu X, Frisina RD. Gene expression changes for antioxidants pathways in the mouse cochlea: relations to age-related hearing deficits. *PLoS ONE.* (2014) 9:e90279. doi: 10.1371/journal.pone.0090279
5. Zheng Z, Tang D, Zhao L, Li W, Han J, Hu B, et al. Liproxstatin-1 protects hair cell-like HEI-OC1 cells and cochlear hair cells against neomycin ototoxicity. *Oxid Med Cell Longev.* (2020) 2020:1782659. doi: 10.1155/2020/1782659
6. Yang Q, Sun G, Yin H, Li H, Cao Z, Wang J, et al. PINK1 protects auditory hair cells and spiral ganglion neurons from cisplatin-induced ototoxicity via inducing autophagy and inhibiting JNK signaling pathway. *Free Radic Biol Med.* (2018) 120:342–55. doi: 10.1016/j.freeradbiomed.2018.02.025
7. He Y, Cai C, Sun S, Wang X, Li W, Li H. Effect of JNK inhibitor SP600125 on hair cell regeneration in zebrafish (Danio rerio) larvae. *Oncotarget.* (2016) 7:51640–50. doi: 10.18632/oncotarget.10540
8. Yin H, Sun G, Yang Q, Chen C, Qi Q, Wang H, et al. NLRX1 accelerates cisplatin-induced ototoxity in HEI-OC1 cells via promoting generation of ROS and activation of JNK signaling pathway. *Sci Rep.* (2017) 7:44311. doi: 10.1038/srep44311
9. Dolasia K, Bisht MK, Pradhan G, Udgata A, Mukhopadhyay S. TLRs/NLRs: shaping the landscape of host immunity. *Int Rev Immunol.* (2018) 37:3–19. doi: 10.1080/08830185.2017.1397656
10. Zhang H, Xiao Y, Nederlof R, Bakker D, Zhang P, Girardin SE, et al. NLRX1 deletion increases ischemia-reperfusion damage and activates glucose metabolism in mouse heart. *Front Immunol.* (2020) 11:591815. doi: 10.3389/fimmu.2020.591815
11. Tattoli I, Carneiro LA, Jéhanno M, Magalhaes JG, Shu Y, Philpott DJ, et al. NLRX1 is a mitochondrial NOD-like receptor that amplifies NF-kappaB and JNK pathways by inducing reactive oxygen species production. *EMBO Rep.* (2008) 9:293–300. doi: 10.1038/sj.embor.7401161
12. Singh K, Roy M, Prajapati P, Lipatova A, Sripada L, Gohel D, et al. NLRX1 regulates TNF-α-induced mitochondria-lysosomal crosstalk to maintain the invasive and metastatic potential of breast cancer cells. *Biochim Biophys Acta Mol Basis Dis.* (2019) 1865:1460–76. doi: 10.1016/j.bbadis.2019.02.018
13. Pickering RJ, Booty LM, NLR. in eXile: Emerging roles of NLRX1 in immunity and human disease. *Immunology.* (2021) 162:268–80. doi: 10.1111/imm.13291
14. Singh K, Poteryakhina A, Zheltukhin A, Bhatelia K, Prajapati P, Sripada L, et al. NLRX1 acts as tumor suppressor by regulating TNF-α induced apoptosis and metabolism in cancer cells. *Biochim Biophys Acta.* (2015) 1853:1073–86. doi: 10.1016/j.bbamcr.2015.01.016
15. Aziz A, Md Daud MK, Nik Othman NA, Abd Rahman N. Early detection of high-frequency presbycusis among normal hearing individuals. *Otol Neurotol.* (2020) 41:e989–92. doi: 10.1097/MAO.0000000000002725
16. Ohlemiller KK. Mechanisms and genes in human strial presbycusis from animal models. *Brain Res.* (2009) 1277:70–83. doi: 10.1016/j.brainres.2009.02.079

17. Suzuki J, Inada H, Han C, Kim MJ, Kimura R, Takata Y, et al. "Passenger gene" problem in transgenic C57BL/6 mice used in hearing research. *Neurosci Res.* (2020) 158:6–15. doi: 10.1016/j.neures.2019.10.007
18. Arnoult D, Soares F, Tattoli I, Castanier C, Philpott DJ, Girardin SE. An N-terminal addressing sequence targets NLRX1 to the mitochondrial matrix. *J Cell Sci.* (2009) 122:3161–8. doi: 10.1242/jcs.051193
19. Kastelberg B, Tubau-Juni N, Ayubi T, Leung A, Leber A, Hontecillas R, et al. NLRX1 is a key regulator of immune signaling during invasive pulmonary aspergillosis. *PLoS Pathog.* (2020) 16:e1008854. doi: 10.1371/journal.ppat.1008854
20. Guo M, Lv M, Shao Y, Zhang W, Zhao X, Li C. Bax functions as coelomocyte apoptosis regulator in the sea cucumber Apostichopus japonicus. *Dev Comp Immunol.* (2020) 102:103490. doi: 10.1016/j.dci.2019.103490
21. Laulier C, Lopez BS. The secret life of Bcl-2: apoptosis-independent inhibition of DNA repair by Bcl-2 family members. *Mutat Res.* (2012) 751:247–57. doi: 10.1016/j.mrrev.2012.05.002
22. Zhang Y, Yang X, Ge X, Zhang F. Puerarin attenuates neurological deficits via Bcl-2/Bax/cleaved caspase-3 and Sirt3/SOD2 apoptotic pathways in subarachnoid hemorrhage mice. *Biomed Pharmacother.* (2019) 109:726–33. doi: 10.1016/j.biopha.2018.10.161
23. Jeng JY, Johnson SL, Carlton AJ, De Tomasi L, Goodyear RJ, De Faveri F, et al. Age-related changes in the biophysical and morphological characteristics of mouse cochlear outer hair cells. *J Physiol.* (2020) 598:3891–910. doi: 10.1113/JP279795
24. Ma D, Zhao Y, She J, Zhu Y, Zhao Y, Liu L, et al. NLRX1 alleviates lipopolysaccharide-induced apoptosis and inflammation in chondrocytes by suppressing the activation of NF-κB signaling. *Int Immunopharmacol.* (2019) 71:7–13. doi: 10.1016/j.intimp.2019.03.001
25. Yang Q, Sun G, Cao Z, Yin H, Qi Q, Wang J, et al. The expression of NLRX1 in C57BL/6 mice cochlear hair cells: Possible relation to aging- and neomycin-induced deafness. *Neurosci Lett.* (2016) 616:138–46. doi: 10.1016/j.neulet.2015.11.053
26. Chen J, Sun D, Shi Y, Jin W, Wang Y, Xi Q, et al. dynamic alterations in spontaneous neural activity in mild cognitive impairment patients with depression symptom: a resting-state fMRI study. *Front Neurosci.* (2019) 13:1333. doi: 10.3389/fnins.2018.00994
27. Sun Y, Ge X, Han X, Cao W, Wang Y, Ding W, et al. Characterizing brain iron deposition in patients with subcortical vascular mild cognitive impairment using quantitative susceptibility mapping: a potential biomarker. *Front Aging Neurosci.* (2017) 9:81. doi: 10.3389/fnagi.2017.00081
28. Anter AM, Wei Y, Su J, Yuan Y, Lei B, Duan G, et al. A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI. *Inf Sci.* (2019) 503:670–87. doi: 10.1016/j.ins.2019.07.026
29. Wang Y, Yang Y, Na S, Liu T, Hang C, Jiang B, et al. Melatonin attenuates early brain injury via regulating miR-181a/TNF-α/NF-κB signaling pathway following subarachnoid hemorrhage in rat. *Acta Med Mediterr.* (2020) 36:3377–83.
30. Li N, Zhang YM, Wang YJ. Abstract TP457: Brain functional alterations in patients with leukoaraiosis associated mild cognitive impairment. *Stroke.* (2016) 47:457–457. doi: 10.1161/str.47.suppl_1.tp457
31. Macpherson H, Formica M, Harris E, Daly RM. Brain functional alterations in type 2 diabetes–a systematic review of fMRI studies. *Front Neuroendocrinol.* (2017) 47:34–46. doi: 10.1016/j.yfrne.2017.07.001
32. Wang J, Chen H, Yuexiu LI, Na YE, Feng L, Wang S, et al. Small-world networks analysis of leukoaraiosis: a rest-state functional MRI study. *Chin J Behav Med Brain Sci.* (2017) 26:977–82.
33. Chen J, Duan X, Shu H, Wang Z, Long Z, Liu D, et al. Differential contributions of subregions of medial temporal lobe to memory system in amnestic mild cognitive impairment: insights from fMRI study. *Sci Rep.* (2016) 6:26148. doi: 10.1038/srep26148

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Li, Bian, Luo, Zhang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 29 March 2022
doi: 10.3389/fpubh.2022.870822

A Study of Family Process Factors of Social Anxiety on the Internet Based on Big Data—Take Guangxi University Students as an Example

*Dexu Bi**

Department of Elementary Education, Guangxi Police College, Nanning, China

Background: Along with the popularization of the new medium of interpersonal communication, many researchers have found that the use of social media has brought about many mental health problems. For example, the virtual nature, vulnerability, and uncertainty of online communication lead to reduced online trust, causing interaction anxiety (IA). The data footprints left on the Internet are processed by malicious elements for big data, leading to the leakage of personal privacy data, bringing content sharing anxiety (SAC) and privacy concern anxiety (PAC), which are all typical forms of online social anxiety. In the face of this situation, analyzing the influence of online social networking on the social psychology of university students and guiding it has become an inevitable issue in the Internet era.

Methods: Learning from the classification of family environment, a self-administered family process factor questionnaire and the Social Anxiety Scale for Social Media Users (SAS-SMU) were used to investigate the online social anxiety of Guangxi University students. The study used SPSS26.0 and Stata for data analysis and descriptive statistics, ANOVA, *t*-test, and linear regression analysis were used to explore the relationship between family process factors and online social anxiety of the university students.

Results: The results showed that except for parental supervision ($p > 0.05$), the effects of interparental relationship, parent-child relationship, sibling relationship, and family atmosphere on university students' online social anxiety were statistically significant and showed positive correlations (F/t = 6.64, 3.53, 4.15, 5.94; $p < 0.05$). Multiple linear regression analysis showed that university students' total online social anxiety score = 36.914−4.09 × good parental relationship−4.16 × good family atmosphere−3.42 × good sibling relationship.

Conclusions: Based on the family systems theory, it is suggested that a comprehensive intervention should be conducted for the coupled system (parental relationship) and sibling system (non-only child's sibling relationship) in the family and focus on the protective factors of parental harmony, sibling relationship harmony, and relaxed family atmosphere. In the specific implementation method, the collaborative shared healthcare plan (CSHCP) can be used to strengthen remote family emotional interaction and avoid

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Praveen Kumar,
VIT University, India
Abdul Rehman Javed,
Air University, Pakistan

***Correspondence:**
Dexu Bi
969224380@qq.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *07 February 2022*
Accepted: *21 February 2022*
Published: *29 March 2022*

Citation:
Bi D (2022) A Study of Family Process Factors of Social Anxiety on the Internet Based on Big Data—Take Guangxi University Students as an Example. Front. Public Health 10:870822. doi: 10.3389/fpubh.2022.870822

Internet addiction. For university students with online social anxiety disorders, their personal health records (PHRs) can be maintained permanently and safely using the Star File System (IPFS), in addition to the convenience of IPFS data extraction, which is more conducive to the timely and long-term tracking treatment of anxious university students.

Keywords: family systems theory, Guangxi University students, online social anxiety, family process factors, big data

INTRODUCTION

With the development of the Internet, there are at least two different forms of modern human interaction activities, namely, real face-to-face interaction in the traditional social presence space and non-face-to-face network interaction in the absence of field cyberspace. As a new form of social interaction, network interaction is changing the way people interact socially and has become an important way of interpersonal communication. The 2020 Digital Report, published by We Are Social, shows that more than 4.5 billion people use the Internet, and social media users have surpassed 3.8 billion (1). According to the 46th Statistical Report on the Development Status of the Internet in China released by the China Internet Network Information Center (CNNIC), Chinese Internet users reached 940 million by June 2020. The scale of instant messaging users reached 99.0% of the overall Internet users (2). It can be said that socializing through the Internet has become the most important form of human interaction.

Along with the popularization of the new medium of interpersonal communication, many researchers have found that the use of social media has brought about many mental health problems. For example, the virtual nature, vulnerability, and uncertainty of online communication lead to reduced online trust, causing interaction anxiety (IA). Social networking data is maliciously collected and leaked by people with the help of big data technology, resulting in content sharing anxiety (SAC) and privacy concern anxiety (PAC) among network users. Online social anxiety is a significant and persistent emotional response of worry, fear, and avoidance behavior to one or more interpersonal situations in the interpersonal exchange of information and emotions based on the Internet and mediated by computers (3). Basic performance behaviors are fear of interacting with others, fear of embarrassing oneself with the content shared, fear of chatting with others online, and fear of others making negative comments about oneself because of the content shared, etc.

As the main group of social media users, university students are troubled by "uncertainty," "social addiction," and privacy leakage while gaining access to various convenient services brought by the Internet, thus, giving rise to online social anxiety. In the face of this situation, it has become an inevitable problem to analyze the influencing factors of university students' online social networking and guide it. As a result, this article will select university students in Guangxi as the research object and take family process factors as the link to empirically analyze the family process factors of online social anxiety in university students in order to provide guidance for the treatment and behavioral improvement of online social anxiety.

Progress of Research on Social Anxiety in Adolescents Guided by Family Systems Theory

American scholar M. Bowen and his collaborator M.E. Kerr's family systems theory (4, 5) considers the family as a "grouped emotional body" that is the result of the interaction between the "father-mother-child" triangle in which individuality or separateness and togetherness or fusion are intertwined—individuality (or separateness) and togetherness (or fusion). Individuality and togetherness are two important balancing agents in the family system. The former seeks independence, while the latter seeks family belonging. Ideally, the family relationship is one in which these two forces are in balance. In this case, family members have their own independence and maintain close ties with each other. Suppose the family relationship is too close and the children are too emotionally involved with their families. In that case, the children will have low levels of Emotional Fusion and Differentiation and will be less independent. Conversely, if they are too distant, they have high levels of Emotional Fusion and Differentiation and are too independent. Both of these imbalance models produce dysregulation of the nuclear family emotional management system, where children need to pay constant attention to their parents' demands and emotional performance, are unable to think and act according to their own needs, are under high tension for long periods of time, have excessive emotional stress intensity and become increasingly attached to becoming independent, resulting in interpersonal difficulties within the family. Suppose individuals respond to the emotional needs of others by choosing to cluster or choosing Emotional Fusion and Differentiation in a fragmented way. In that case, they will develop anxiety and show anxious behavior.

Bowen's family systems theory has provided new research ideas for adolescent clinical psychology research. Based on this theory, many scholars have examined the mechanisms by which family systems interact with each other in the psychogenesis and development of adolescents, especially providing a new analytical framework for exploring the influence of parent-child relationship quality on the psychosocial adjustment of family members. Drawing on Bowen's family systems theory, Williams (6) further generalized the family environment into non-process and process factors based on the presence or absence of family members' interactions. The non-process factors refer to the members' own characteristics or inherent characteristics of the

family environment, which mainly include static factors, such as family structure, family socioeconomic status, primary provider, primary caregiver, and parents' own characteristics (7). Process factors refer to the interactions between members within the family system and the family climate formed by the interactions that mainly include the couple relationship, parenting, parent-child relationship, non-only child sibling relationship, and the family climate formed by them.

Chinese and foreign scholars have conducted a large number of studies on the effects of family non-process factors on social anxiety. They mainly involve the effects of gender (8, 9), family structure (10), grade level (11), and family economic and social status (12, 13) on social anxiety. The influence on family process factors involves parenting style (14, 15), family closeness (16), family function (17, 18), and parental conflict (19–22). Compared to the non-process factors that influence social anxiety, which are difficult to change, the process factors can be effectively guided and have more plastic value for the benign development of social anxiety. However, from the general status quo of research on family factors of social anxiety at this stage, family process factors are either done as single-factor research, which is in-depth but lacks systematicity, or put into family factors as grand unified research, which does not make specialized distinctions and lacks in-depth insight. Specifically, no directly relevant domestic and international literature has been retrieved on the empirical study of helping university students overcome online social anxiety through the improvement of family process factors.

MATERIALS AND METHODS

Study Participants

From September 2021 to October 2021, Guangxi Police College, Guangxi Foreign Language Institute, Nanning College, and Guangxi International Business Vocational and Technical College were selected. The anonymous questionnaire test was conducted with the informed consent of the subjects, and the information of the subjects was kept strictly confidential. A total of 2,400 questionnaires were distributed, and 2,373 valid questionnaires were returned, with a valid recovery rate of 94.6%. The age range was 18–22 years old, with 1,125 men (47.4%) and 1,248 women (52.6%). There were 995 people from rural areas, accounting for 41.9%, and 1,378 people from urban areas, accounting for 58.1%. As for the structure of majors, 1,350 students were studying literature, history, and finance, and 1,023 students were studying science and agriculture, accounting for 56.9 and 43.1%, respectively. In terms of grade structure, the numbers of freshmen, sophomores, juniors, and seniors were 503, 657, 797, and 416, accounting for 21.2, 27.7, 33.6, and 17.5%.

Research Method

The online social anxiety scale was adopted from the Social Anxiety Scale for Social Media Users (SAS-SMU) revised by Wang (23), which was validated and revised based on the online SAS-SMU established by Alkis et al. in Turkey. It consists of three dimensions with 15 entries, and the three dimensions are SAC, PAC, and IA. The 5-point Likert scale was used (not at all, rarely, sometimes, often, and completely), with higher scores indicating a more severe degree of online social anxiety (24). The Cronbach $\alpha = 0.928$ in this survey.

Family process factors include (1) parent-child relationships (poor, fair, and good), and problem behavior theory suggests that a good parent-child relationship will increase their regular behaviors with a corresponding decrease in problem behaviors. (2) Interparental relationships (poor, fair, and good), family dynamics theory suggests that if an individual is in a deteriorating family relationship, it will lead to poor socialization of the child (25). (3) Family climate (poor, fair, and good), research has shown that the more harmonious the family climate is perceived by adolescents, the less they engage in problem behaviors. (4) Parental supervision (rarely, occasionally, and often), according to Jessor's problem behavior theory, the higher the level of support and control (especially from parents) perceived by adolescents, the less likely they are to engage in problem behaviors (26). (5) Sibling relationship (poor, fair, and good).

Data Collection

The survey was conducted by the Public Foundation Department of Guangxi Police College and some students used the actual test. Before the survey, the investigators were trained in a unified manner. In the survey, arranged quality control personnel in the field are responsible for answering possible problems to the survey respondents in a timely manner. The surveyors are responsible for recovering and reviewing the completeness of the completed questionnaires (27). The study used SPSS 26.0 and Stata for data analysis; descriptive statistics, one-way ANOVA, *t*-test, and linear regression analysis were used to explore the relationship between family process factors and online social anxiety of university students. $P < 0.05$ indicates that the difference is statistically significant.

STATISTICAL ANALYSIS

Descriptive Analysis of Family Process Factors of Guangxi University Students

There were 322 (13.6%) only children and 2,051 (86.4%) non-only children among 2,373 university students. The numbers of non-only children's families with poor, fair, and good sibling relationships were 29 (1.41%), 317 (15.5%), and 1,720 (83.1%). The numbers of poor, moderate, and good parental relationships were 161 (6.78%), 640 (27%), and 1,572 (66.25%). The frequency of family supervision was rarely, occasionally, and often for 72 (3%), 1,520 (64.1%), and 781 (32.9%). Those with poor, average, and good parent-child relationships were 31 (1.3%), 692 (29.2%), and 1,650 (69.5%). Poor, fair, and good family atmospheres occurred in 531 (22.4%), 1,330 (56%), and 512 (21.6%).

Differential Analysis of Family Process Factors of Online Social Anxiety

Table 1 shows, except for parental supervision ($p > 0.05$), that the effects of interparental relationship, parent-child relationship, sibling relationship, and family atmosphere on online social anxiety in rural university students were statistically significant and showed positive correlations ($F/t = 6.64, 3.53, 4.15, 5.94$, $p < 0.05$).

TABLE 1 | Differential analysis of family process factors of online social anxiety ($n = 2,373$).

		N	$\bar{x} \pm s$	F/T	*P*
Only child	Yes	322	30.19 ± 10.95	−0.38	0.702
	No	2,051	30.93 ± 10.04		
Relationship with compatriots	Poor	29	42.67 ± 6.43	4.15	0.007
	Fair	317	35.13 ± 10.63		
	Good	1,705	29.72 ± 9.82		
Parent-child relationship	Poor	31	38 ± 17.35	3.53	0.031
	Fair	692	33.12 ± 8.85		
	Good	1,650	29.74 ± 10.38		
Family atmosphere	Poor	531	33.17 ± 10.54	5.94	0.003
	Fair	1,330	31.44 ± 9.13		
	Good	512	26.78 ± 11.27		
Parental supervision	Rarely	72	24.43 ± 10.72	2.58	0.078
	Occasionally	1,520	31.74 ± 9.37		
	Often	781	29.63 ± 11.31		
Inter-parental relations	Poor	161	35.13 ± 12.12	6.64	0.002
	Fair	640	33.83 ± 8.92		
	Good	1,572	29.17 ± 10.06		

Multiple Linear Regression Analysis of Online Social Anxiety

The total score of the online social anxiety scale of university students was used as the dependent variable. The five family process factors of parental relationship, parental supervision, parent-child relationship, sibling relationship, and family atmosphere were used as independent variables for multiple linear regression analysis. First, the covariance test was conducted, the VIF values were all <10, and there was no covariance between the variables. Total anxiety score = 36.914−4.09 × good parental relationship−4.16 × good family atmosphere−3.42 × good sibling relationship, i.e., the total anxiety score was on average 4.09 points lower when comparing good and poor parental relationships. The average total anxiety score was 4.16 points lower when comparing good and poor family atmospheres. The average total anxiety score was 3.42 points lower when comparing good and poor sibling relationships (refer to **Table 2**).

DISCUSSION AND SUGGESTIONS

Risk Factors and Protective Factors in Family Process Factors Affecting Online Social Anxiety of University Students

The results of the study showed that parental relationship, family atmosphere, and sibling relationship influenced Guangxi University students' online social anxiety, where parental friendship, sibling friendliness, and family harmony were protective factors for Guangxi University students. The opposite were risk factors. That is, the more parents punish and control when educating their children, the stronger the domestic violence environment, and the more family conflicts, the higher the incidence of university students' online social anxiety. On the contrary, good parental and sibling relationships and high intimacy in the family are protective factors to reduce the incidence of Guangxi University students' online social anxiety.

The risk factors among the family process factors that triggered or increased the online social anxiety of university students were poor parental relationships, poor sibling relationships in non-one-child families, and overall family climate disharmony. From the viewpoint of family systems theory, these risk factors occur in the coupled system (high conflict between spouses and strong violent environment) and sibling system (unfriendly sibling relationship), leading to children's overall perception of poor family climate playing a role in online social anxiety of university students' promoting effect. It is clear from this that the risky effects of negative family environments occurring in multiple family subsystems are significant, and it is evident that online social anxiety of university students is not the result of the action of a single system in the family but maybe the result of multiple risk systems acting together.

Suggestions From the Perspective of Family Systems Theory

According to stress-coping theory, one of the main reasons that adolescents face a variety of risk factors during their development and only a minority of them end up with problems is that these risk factors do not act on the same adolescent at the same time. Adolescents' own psychological abilities (e.g., mental toughness, high self-efficacy, etc.) counteract the negative consequences of negative factors, a principle that is also valid for online social anxiety. If an adolescent grows up with only one risk factor, such as poor parental relationship

TABLE 2 | Multiple linear regression analysis of online social anxiety ($n = 2,373$).

Dependent variable: Online social anxiety total score of Guangxi university students					
	Unstandardized coefficient		Standardization coefficient	t	Sig
	B	Standard error	Beta		
(Constant)	36.914	1.446		25.521	0
Good parental relationship	−4.094	1.347	−0.191	−3.04	0.003
Good family atmosphere	−4.158	1.548	−0.169	−2.686	0.008
Good sibling relationship	−3.417	1.413	−0.151	−2.418	0.016

or weak parental supervision, or sibling tension, the likelihood of online social anxiety in adolescents is low. However, when many risk factors, such as poor parental relationship, sibling disharmony, and family tension, act on adolescents at the same time, there is a high probability that online social anxiety will occur (28).

Developmental psychopathology suggests that the development of social adjustment problems in individuals is influenced by several factors that increase the risk of developing mental illness, while some factors protect individuals from physical and mental illness. Since the occurrence of online social anxiety is the result of a combination of risk factors, the more risk factors there are for online social anxiety interventions, the more protective factors there are to counteract the risk factors. The results of this empirical analysis of the factors influencing the online social anxiety of Guangxi University students showed that the total anxiety score of good parental relationships was on average 4.09 points lower than that of poor parental relationships. The average anxiety score of a good family atmosphere was 4.16 points lower than that of a poor family atmosphere. The average anxiety score of a good sibling relationship was 3.42 points lower than that of a poor sibling relationship, which were the protective factors of family process factors in reducing online social anxiety. A healthy developing adolescent has a positive support system, while a poorly developing adolescent has a negative support system behind him or her. In addition, there is a pattern of "good together and bad together" between the systems.

Although online social networking makes up for the deficiency of public real interpersonal communication to a certain extent. However, due to the virtuality, vulnerability, and uncertainty of the network, coupled with the proliferation of information, interpersonal overload, and privacy security of online social media, online social networking is mostly superficial and lacks depth, and personal emotional needs cannot be met, which promotes the emergence of online social anxiety and indulges it. As the most important place for emotional breeding and cultivation, the family should become a warm home to avoid and overcome the online social anxiety of university students. Based on the empirical analysis of this thesis, to intervene in the online social anxiety of university students, on the one hand, we need to intervene in the support system behind university students comprehensively. In terms of family process factors, we need to intervene in the coupled system, parent-child system, and sibling system in the family as a whole and improve the education and guidance mechanism of protective factors. On the other hand, when we intervene, we can also seize the key points and key links in the system. In the specific implementation method of correcting network anxiety through family, a generalized collaborative framework named collaborative shared healthcare plan (CSHCP) gives us a great idea to correct online social anxiety of university students (29). It is a CSHCP framework for daily life activity recognition and can be used for remote coordination and communication between family members to enhance emotion and avoid addiction to the network. This is more suitable for college students who leave their families to study. For college students with online social anxiety, their personal health records (PHRs) can be maintained using the Star File System (IPFS) (30), and they can access their records online and make targeted corrections at any time.

CONCLUSION

Family is not only an extremely important environment in the process of individual psychological development and even the formation of psychological quality but also the main influencing factor of individual personality characteristics and mental health. Based on this, the thesis makes some tentative enquiries into the relationship between family process factors and online social anxiety. Although the statistical value of the sample survey cannot infer the overall trend quantitatively, the survey results of the sample can explain some overall characteristics: that parental relationship, family atmosphere, and sibling relationship influenced online social anxiety of university students. To intervene in the online social anxiety of university students, on the one hand, we need to intervene in the coupled system, parent-child system, and sibling system in the family as a whole and improve the education and guidance mechanism of protective factors. On the other hand, we can also concentrate on the key points and key links in the family process system. Because this study is affected by the randomness of the sampling survey and the limitation of the number of questionnaires, and the understanding of online social anxiety will be more or less subjective, the conclusion of the study is not suitable for popularization and needs further verification. Finally, it is hoped that through the relevant elaboration of the family

factors of online social anxiety, the public can pay attention to family harmony.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Wei Wang, Wen Wu, Ling Zhang, Yun Lu, Qing Wang, Guangxi Police College. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

DB provided the overall idea of the article, the project design, and data analysis.

FUNDING

This article titled Research on Social Integration of Migrant Children Relocated to Poverty Alleviation Areas in Guangxi was supported by the research project of Guangxi Philosophy and social sciences in the year 2021 (21FRK003). This article is part of the research results of the research project.

REFERENCES

1. *We Are Social: Global Network Overview Report 2020 [R/OL]*. Available online at: http://www.199it.com/archives/1020700.html (accessed March 18, 2020).
2. China Internet Information Office. *The 46th Statistical Report on the Development Status of China's Internet (Full Text) [R/OL]*. Available online at: http://www.cac.gov.cn/2020-09/29/c_1602939918747816.htm (accessed March 10, 2019).
3. Zhang Y. *Media Dependence, Social Anxiety, and Self-Presentation in the Online Interaction-an Analysis of Online Interaction Behavior of University Students*. Jilin: Jilin University (2019). p. 33.
4. Bowen M. *Family Therapy in Clinical Practice* (1978).
5. Kerr M, Bowen. *Family Evaluation* (1988).
6. Williams NC. *The Relationship of Home Environment and Kindergarten Readiness. A Dissertation for the Degree Doctor in Education*. East Tennesee State University (2002). p. 15–35.
7. He Y. Analysis of social anxiety and its influencing factors among left-behind children in rural areas. *Modern Vocat Educ*. (2021) 40:32–3.
8. Wu W. The relationship between social anxiety and professional competence determination and perceived social support among specialized medical students. *Psychol Month*. (2021) 16:36–8. doi: 10.16835/j.cnki.1000-9817.2021.16.022
9. Liu G. The influence of adolescents' psychological quality on social anxiety and its gender differences. *J Southw Normal Univ*. (2019) 44:91–6. doi: 10.13718/j.cnki.xsxb.2019.02.016
10. Wu M. *The Effect of Secondary School Girls' Family Parenting Style on Social Anxiety -A Study on the Mechanism of Action and Intervention of Explanation Bias [D]*. Jiangxi: Jiangxi Normal University (2020). p. 45–6.
11. She X. *Study on the Relationship Between Family Parenting Style, Emotion Regulation and Social Anxiety and Intervention Among Middle School Students [D]*. Hubei: Huazhong Normal University (2020). p. 12–3.
12. Zhao G. *The Relationship Between Secondary School Students' Depressive Mood and Their Family Socioeconomic Status and Mental Toughness [D]*. Guizhou: Guizhou Normal University (2020). p. 23–4.
13. Li X, Ren Z, Hu X, Guo Y. Why are low family social class university students more prone to social anxiety? –Multiple mediating roles of psychosocial resources and rejection sensitivity. *Psychol Sci*. (2019) 42:1354–60. doi: 10.16719/j.cnki.1671-6981.20190611
14. Fox JK, Masia Warner C. Assessing clinical improvement in school-based treatment for social anxiety disorder: agreement between adolescents, parents, and independent evaluators. *Child Psychiatry Hum Dev*. (2017) 48:721–7. doi: 10.1007/s10578-016-0697-5
15. Wang B, Cao J, Zeng C. The mediating role of self-acceptance and parenting style in moderating the relationship between gender roles and social anxiety in middle and high school students. *Chinese J Child Health*. (2021) 29:829–33. doi: 10.14045/j.cnki.rhen.2021.08.007
16. Wang X. *A Study on the Relationship Between Family Closeness and Social Anxiety Among Elementary School Students - Mediating Effects of Self-Esteem and Social Self-Perception [D]*. Hebei: Hebei Normal University (2020). p. 22–3.
17. Wang Y, Li Y, Wu F. The relationship between family functioning and preschool children's behavioral problems: the multiple mediating roles of attachment avoidance and online social anxiety. *Dev Psychol Educ*. (2021) 37:76–83. doi: 10.16187/j.cnki.issn1001-4918.2021.01.10
18. Liu X, Li Y, Xie Q, Wu F, Suo H. The relationship between maternal attachment anxiety and preschoolers' online social anxiety: the mediating chain role of marital conflict and parent-child attachment. *Chinese J Clin Psychol*. (2020) 28:700–4. doi: 10.16128/j.cnki.1005-3611.2020.04.010
19. Meng L. *The Effect of Parental Conflict on Social Anxiety in University Students: The Mediating Role of Intimate Fear and the Moderating Role of Interpersonal Trust*. Liaoning: Shenyang Normal University (2021). p. 16–17.
20. Jiang N. *A Follow-Up Study on the Relationship Between Parental Conflict and Mid- to Late-Childhood Aggression: The Mediating Role of Social Anxiety*. Shandong: Shandong Normal University (2021). p. 22.
21. Li S. *The Relationship Between Parental Marital Conflict and Adolescents' Subjective Well-Being-A Study on the Role of Parent-Child Attachment and Emotional Security and Intervention*. Hubei: Huazhong Normal University (2019).
22. Weymouth BB, Fosco GM, Mak HW, Mayfield K, LoBraico EJ, Feinberg ME. Implications of interparental conflict for adolescents' peer relationships: a longitudinal pathway through threat appraisals and social anxiety symptoms. *Dev Psychol*. (2019) 55:1509–22. doi: 10.1037/dev0000731
23. Wang Y. A test of social anxiety scale for social media users [J]. *Young J*. (2021) 50–1. doi: 10.15997/j.cnki.qnjz.2021.06.019
24. Sun M, Liu K, Liu L, Meng L, Huang L. Reliability evaluation of the streamlined version of the adolescent social anxiety scale in a population of junior high school students [J]. *Modern Prevent Med*. (2017) 23:4310–3.
25. Wang H, Cui M. Research on the influence of family factors in the formation of individual gender roles. *J Northeast Normal Univer*. (2014) 171–4. doi: 10.16164/j.cnki.22-1062/c.2014.02.051
26. Zhang X, Wang L. Family and juvenile delinquency. *Modern Enterprise Educ*. (2007) 45–6.
27. Bi D, Chang L, Wang Q, Wu W. Family correlates of physical aggression among students in concentrated poverty-stricken minority areas. *Chinese J School Health*. (2019) 40:239–41. doi: 10.16835/j.cnki.1000-9817.2019.02.022

28. Dong H. *Research on Family Coping Strategies for Child Bullying*. Beijing: People's Education Press (2020). p. 188–9.
29. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfik H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. *Hum Cent Comput Inf Sci.* (2020) 10:40. doi: 10.1186/s13673-020-00245-7
30. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an IPFS-Based optimization algorithm. *J Circuits Syst Comput*. (2022) 31:2250010. doi: 10.1142/S0218126622500104

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 30 March 2022
doi: 10.3389/fpubh.2022.855680

The Text Mining Technique Applied to the Analysis of Health Interventions to Combat Congenital Syphilis in Brazil: The Case of the "Syphilis No!" Project

Marcella A. da Rocha[1], Marquiony M. dos Santos[1], Raphael S. Fontes[1], Andréa S. P. de Melo[1], Aliete Cunha-Oliveira[2], Angélica E. Miranda[3], Carlos A. P. de Oliveira[4], Hugo Gonçalo Oliveira[5], Cristine M. G. Gusmão[6], Thaísa G. F. M. S. Lima[7], Rafael Pinto[1], Daniele M. S. Barros[1] and Ricardo A. de M. Valentim[1]*

[1] Laboratory of Technological Innovation in Health, Federal University of Rio Grande do Norte, Natal, Brazil, [2] Health Sciences Research Unit: Nursing (UICISA:E), Nursing School of Coimbra (ESEnfC), Coimbra, Portugal, [3] Postgraduate Program in Infectious Diseases, Federal University of Espírito Santo, Vitoria, Brazil, [4] Multidisciplinary Department of Human Development With Technologies, State University of Rio de Janeiro, Rio de Janeiro, Brazil, [5] Centre for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal, [6] Department of Biomedical Engineering, Federal University of Pernambuco (UFPE), Recife, Brazil, [7] Ministry of Health, Brasília, Brazil

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Abdul Rehman Javed, Air University, Pakistan

***Correspondence:**
Daniele M. S. Barros
daniele.barros@lais.huol.ufrn.br

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *15 January 2022*
Accepted: *04 February 2022*
Published: *30 March 2022*

Citation:
da Rocha MA, dos Santos MM, Fontes RS, de Melo ASP, Cunha-Oliveira A, Miranda AE, de Oliveira CAP, Oliveira HG, Gusmão CMG, Lima TGFMS, Pinto R, Barros DMS and Valentim RAM (2022) The Text Mining Technique Applied to the Analysis of Health Interventions to Combat Congenital Syphilis in Brazil: The Case of the "Syphilis No!" Project. Front. Public Health 10:855680. doi: 10.3389/fpubh.2022.855680

Congenital syphilis (CS) remains a threat to public health worldwide, especially in developing countries. To mitigate the impacts of the CS epidemic, the Brazilian government has developed a national intervention project called "Syphilis No." Thus, among its range of actions is the production of thousands of writings featuring the experiences of research and intervention supporters (RIS) of the project, called field researchers. In addition, this large volume of base data was subjected to analysis through data mining, which may contribute to better strategies for combating syphilis. Natural language processing is a form of knowledge extraction. First, the database extracted from the "LUES Platform" with 4,874 documents between 2018 and 2020 was employed. This was followed by text preprocessing, selecting texts referring to the field researchers' reports for analysis. Finally, for analyzing the documents, N-grams extraction (N = 2,3,4) was performed. The combination of the TF-IDF metric with the BoW algorithm was applied to assess terms' importance and frequency and text clustering. In total, 1019 field activity reports were mined. Word extraction from the text mining method set out the following guiding axioms from the bigrams: "confronting syphilis in primary health care;" "investigation committee for congenital syphilis in the territory;" "municipal plan for monitoring and investigating syphilis cases through health surveillance;" "women's healthcare networks for syphilis in pregnant;" "diagnosis and treatment with a focus on rapid testing." Text mining may serve public health research subjects when used

in parallel with the conventional content analysis method. The computational method extracted intervention activities from field researchers, also providing inferences on how the strategies of the "Syphilis No" Project influenced the decrease in congenital syphilis cases in the territory.

Keywords: text mining, public health, congenital syphilis, content analysis, "no syphilis" project

1. INTRODUCTION

Syphilis is a sexually transmitted disease (STD) that surfaced in Europe in the late 15th-century (1, 2). While this condition is preventable, curable, and usually treatable with penicillin injections, it remains a global health security threat. In addition, syphilis is the second major cause of adverse pregnancy outcomes, including neonatal death (3–6). That fact raises the syphilis epidemic as a neglected problem that poses considerable challenges to many health systems.

Although there is an effective therapy to combat the infectious agent, Treponema pallidum, measures to control and prevent the disease require that health care systems integrate many levels of care. That includes planning and deploying preventive and therapeutic interventions for all stages and types of syphilis in different population groups. In sum, the infection is relatively simple to treat if diagnosed at an early stage but difficult to control among populations (3, 4).

In 2016, the World Health Organization (WHO) launched an initiative to eliminate congenital syphilis (CS). Its goal was that CS case counts would not exceed 50 cases per 100,000 live births in 80% of countries (7). However, the overall rate was 473 cases per 100,000 live births in this very period (8). In Brazil, rates of reported CS outnumber those of other nations, with 880 cases per 100,000 live births in 2017, thus representing an Average Annual Percent Change (AAPC) of 60.38% (8).

Confronted by the growing syphilis epidemic, the Ministry of Health (MoH) and governing bodies of Brazil's Unified Health System (SUS) developed a national project for implementing and integrating a syphilis response into healthcare networks (9). The "Syphilis No!" Project (SNP), as it was titled, represented the instrument with which the MoH implemented a public health response to reduce syphilis in pregnant women (SIP) and eliminate congenital syphilis in Brazil. It has been in progress since 2018 and is built on two cornerstones: (1) actions with a universal scope and (2) specific actions in priority areas. Specific actions were developed in 100 municipalities the Ministry of Health considered a priority due to the high burden of CS (9, 10).

A team of field researchers carried out the specific actions in line with the five thematic areas of the project (Surveillance, Management and Governance, Integral Care, Education, and Communication) to qualify the local response on syphilis. It seems clear that describing the work of RIS in the priority areas renders substantive material to support analyses on the impact of interventions and their associations as a tool to induce health policy responses to syphilis.

The "LUES Platform" was one of the resources of the project. It was developed and used to monitor field researchers' activities. In such a digital platform, the researchers entered reports or individual texts related to their activities in the territory. Nonetheless, well beyond the need to interpret data on the project's associations in the territory, it is clear that research approaches that interpret textual production (e.g., content or thematic analysis) pose challenges. This is closely related to the scarcity of consolidated methods for using and processing large volumes of text.

Theorists like Bardin (2011) have discussed computers' relevance and practicality to deepen content analysis techniques, as long as researchers prepare unambiguous instructions. From this standpoint, content analysis automation can occur in varying degrees (total or partial). That implies direct consequences on the analytic practice, such as increased speed, increased rigor in research setup, and the potential for complex data manipulation, and so on (11).

An additional challenge for research done with sheer volumes of text data is to objectively interpret structures that aid in understanding the impact of work included on reports available on digital platforms, or at least to attempt to do so. Thus, it would take years to analyze such types of data through conventional manual methods. To address this issue, there are software-related strategies for text analysis that are typically applied in the healthcare field (12).

However, the currently available software is insufficient for monitoring strategic intervention activities. This mostly happens because they produce limited information that does not consider textual modeling (13). Such software limits the ability of researchers to handle the results generated safely. In addition, there are several difficulties in interpreting the word processing in such programs (13).

As the use of Artificial Intelligence continues to expand across many healthcare fields, machine learning algorithms (MLA) for big data analysis and modeling started to emerge (14, 15). In this vein, the text mining technique has the potential to provide consistency in the interpretation of text monitoring strategies, as well as a reliable treatment for the rendering of word form n-grams (16). However, the strategy of using text mining techniques to identify the main intervention activities in large national projects, such as the "Syphilis No!" Project, has not been found within the current literature.

Incorporating text and data mining (TDM) and the content analysis method into research involving large volumes of data has the potential to support evaluative studies (17). Further, it can supplement indicators that assess the effectiveness of work processes and their performance, which will further assist public health strategies toward interventions dealing with challenges similar to congenital syphilis. Therefore, with the aid of a set of

algorithms – which make up a computational method applied to text mining – it is possible to identify which intervention activities developed by the RIS may have spurred public health policy strategies in the territory. In other words, strategies that point to a significant reduction in CS cases.

This article considers that the "Syphilis No!" Project has been contributing to syphilis cases reduction in Brazil. Our assumption is that field researchers' practices demonstrate the link between the goals of the project and the management of syphilis in the territory. Furthermore, their activities could have potentially influenced the indicators of syphilis in priority municipalities.

Hence, this article's scope is to analyze the connections amid the goals of the "Syphilis No!" Project and the content presented by the field researchers by using computational methods that apply text mining algorithms–in addition to discussing the role of such connections as health public policy drivers.

1.1. Related Works

We consider it adequate to briefly quote some studies that use computational methods, whether NLP methods or techniques based on Machine Learning (ML) to differentiate our use of Natural Language Processing. The analysis aimed to identify applications in several areas related to health. El Kah and colleagues studied a large volume of data produced through the implementation of an electronic patient record system (18). To efficiently handle these complex data, researchers employed NLP to extract patterns and elements of interest. By Valentim et al. (19) studied a similar Stochastic Petri Net Model describing the relationship between reported maternal and congenital syphilis cases in Brazilaspect; however, in a different field of application, i.e., the Virtual Learning Environment of the Brazilian Health System (AVASUS). In that example, the goal was to identify the writing style for authorship recognition using NLP methods. In (20), it was possible to find some methodological similarities regarding the work developed by Rocha et al. (21). Machine Learning methods enabled an optimized and more accurate analysis of text content and regular expressions. In our case, applying neural networks also proved to be a feasible way to work with NLP and TDM techniques.

Accordingly, the Clinical Record Interactive Search (CRIS) is another work applied in healthcare by means of content analysis. This work is developed within the context of large volumes of clinical data provided by the South London and Maudsley Trust. Its focus was to search for highly relevant elements in unstructured text, and for this purpose, NLP proved to be particularly important. Such practice prompted access to relevant results about the clinical status of patients.

Thus, it was observed that nearly 7,500 people affected by schizophrenia presented with a clinical picture of negative symptoms, which often leads to impaired cognitive ability, disordered thinking, and lack of concentration. Thus, NLP approaches were proposed, which enabled obtaining a more assertive context from the documents (22).

Regarding electronic patient records, the high volume of these records always stands out, especially when health information systems are centralized. Therefore, computational methods applied to content analysis can significantly contribute to this field, especially in patients' clinical histories studies.

Some studies focus on ubiquitous data acquisition, such as the work developed by Javed et al. (23). Such a study was underpinned by a generic, intelligent, autonomous, and collaborative Internet of Things (IoT) architecture. The proposed model is based on smartphones and machine learning. One of its most potent contributions is sharing health-related information, which can be used in comparative analysis. Furthermore, the ubiquitous model proposed in this study fosters the centralization and production of large volumes of data. Therefore, computational methods, especially regarding content analysis, can be of much value to measure, both quali- or quantitatively, the behavior of users of collaborative health settings, where the main objective is to evaluate the daily activities of these users.

These research efforts were applied in different contexts in the health field, yet they all used computational methods, whether NLP methods or techniques based on Machine Learning (ML). The studies in question approached aspects of patient care and were applied in clinical settings. However, computational methods were used in different contexts in all the studies, which indicates a fertile ground for scientific research and diverse applications. It is noteworthy that none of the studies listed used Natural Language Processing or Machine Learning to evaluate the impact of public health policies, which is one of the main contributions of this article.

2. MATERIALS AND METHODS

According to Bardin (11), content analysis is a set of methodological tools that, being continuously perfected, can be applied to varying types of text. These tools underscore the development of data categories and ratify how relevant it is to understand the significance of the context in which analyzed items have been applied (11).

The primary function of content analysis is to unravel the critic about a particular discourse (11). It can be done qualitatively or quantitatively. In the first approach, the analysis is performed both systematically and analytically. Hence, a researcher reviews themes or categories of analysis through conceptualization, data collection, analysis, and interpretation. In the second, the search is oriented for the quantification of content as to predetermined categories; and it is conducted in a systematic and replicable manner (11, 24).

In both approaches, the goal is to infer knowledge about the semantics of the prevailing discourse in the data or its message. In this study, we adopted a qualitative-quantitative approach. This stems from the fact that such an approach mixes TDM techniques and a conceptualization of the extracted discourse–with a theoretical basis that seeks to understand the intervention activities–from the perspective of public health in the territory. In this context, the theoretical basis that served as a contributing factor in determining this research's development steps was inspired by Bardin (11), and it is characterized into three main stages, namely: (i) pre-analysis; (ii) exploring the materials:

text mining; and (iii) analysis of Obtained Results: Inference and Interpretation, implemented in the Python programming language version 3.7.

These steps were essential to the development of this work. Given the volume of texts produced by the field researchers, the text mining method and subsequent content analysis were the most suitable approaches to study our research subject. Moreover, text mining was necessary to analyze the formation of the most relevant word statistically' and subsequently form logical semantic sentences based on content analysis.

2.1. Pre-analysis

Pre-analysis is the process in which the corpus is organized so it becomes operational and by which preliminary ideas are systematized. In this step, we carried out three tasks: (i) skimming, for an overall comprehension of the documents that would be collected; (ii) documents selection, which consisted in determining which ones would be analyzed; and (iii) formulating the hypotheses and objectives of the research to be developed (11).

In the pre-analysis, it was necessary to grasp (i) the central features that characterize the "Syphilis No!" Project in order to propose the subject of the present study; (ii) the ethical issues related to the research; and (iii) the features of the corpus (texts) for analysis in the next step—exploring the materials.

2.2. Features of the "Syphilis No!" Project

The SNP was developed by the Brazilian government through the Ministry of Health (MH). Over the last ten years, it has been the leading syphilis-related public health intervention in Brazil (8, 10, 25).

Within the breadth of strategies to combat syphilis, this study focuses on the institutional support network composed of field researchers who worked in priority municipalities in the regions of Brazil. Such a support network was assembled with the primary purpose of strengthening the "nexus between the Project and the health services managers in the territory," "to coordinate programmatic actions agreed upon within governance bodies of SUS with local plans" and to offer support for a "timely response to syphilis within the healthcare networks" (26).

The project relied upon 52 research and intervention supporters distributed across 72 of the 100 priority areas in all five Brazilian regions (26). RIS worked alongside health services managers to find the best strategies to confront syphilis. The following determinations guided them: evaluating the health plans and programs of local health departments; strengthening committees for investigating mother-to-child transmission (MTCT); strengthening strategic information systems for health surveillance and improving case reporting, laboratory follow-up, and closure of syphilis cases, as well as operationalizing the line of care for adult syphilis and congenitally exposed infants (10).

All activities developed by RIS were monitored by supervisors who used a digital management system, that is, the "LUES Platform." Through the platform, the researchers produced and submitted individual monthly reports in free text format (not systematized). Researchers produced and submitted individual monthly reports in free-form (not systematized) through the platform. These reports were based on the on-site experience and endeavors, as well as on the results of interactions between the project's actions and interventions in the public health agenda agreed upon in the inter-federative context of Brazil's SUS. It is particularly noteworthy that throughout the operationalization of the SNP, the Ministry of Health, responsible for implementing health policies in response to syphilis in Brazil, assigned the role of coordinating the syphilis policy and its set of actions at the local level to each RIS (8).

2.3. Research Ethics Issues

The database of texts related to the SNP interventions used in this article is in the public domain. In this manner, secondary data are used without any sensitive identification of the participants or public actors. Data were gathered after the extraction and anonymization of the documents released in the repository of the "LUES Platform." Of note, this data can be easily accessed by any researcher at the following link: http://vigilanciasaude.ufrn.br/files/anonymous_reports.csv.

2.4. Exploring the Materials

Each field researcher individually wrote the reports selected for content analysis. Texts could be written in a free format (not systematized), meaning that each supporter developed their own reports featuring an account of what was done and accomplished. Therefore, the project did not provide any template for RIS to follow. Of note, the MoH entrusted each field researcher with a set of actions to be carried out on a monthly basis during the SNP. Actions were aligned with the project's goals and the public health agenda.

Since reports featured free-form content and the health terms structures are identifiable—using the analytical method, by two, three, or more words—the TDM process was fully automated by text mining algorithms using N-grams (see "Words and data extraction" for detailed information). By applying the analytical method, we observed it was not enough to identify any intervention action taken by the field researchers that only one term. However, from two to five words, identification could be achieved.

Text mining may be defined as the automated discovery of hitherto unknown knowledge utilizing an unstructured database (14, 15). A keystone of this method is linking combined information to form new facts or new hypotheses to be explored.

The following actions were carried out for implementing the text mining technique in this research: (i) detailing the Database; (ii) pre-processing; (iii) Words and Data Extraction; and (iv) N-grams Extraction. Each of the activities is subsequently described in detail.

2.5. Detailing the Database

The database was extracted from the "LUES Platform." The texts collected refer to accounts about technical visits and meetings; and progress reports concerning work and research plans, thus totaling 4,874 text file documents–this sum corresponded to 3.86 Gigabytes of digitally stored files. Each document has nearly 740

TABLE 1 | Document Indexing Data.

Variable	Description
Identifier	Document identification number
Month	Month the document was created
Year	Year the document was created
State	Brazilian State where the document was created
Region	Brazilian Region where the document was created

words. The time frame for compiling the corpus was from May 2018 to December 2020.

In order to build the database, those documents underwent the process of anonymization and identification. Afterward, they were indexed according to the information provided in **Table 1**.

Each entry represents a document that has been collected. This document catalog has been saved in a comma-separated values (CSV) format, a text file separated by commas (27). The file is available for download at the following link: http://vigilanciasaude.ufrn.br/files/anonymous_reports.csv.

2.6. Pre-processing

Pre-processing consisted of 4 tasks: (i) Document Standardization; (ii) Text Normalization; (iii) Removal of Duplicate Documents; and (iv) Text Correction. The objective of the first task was to standardize the text database by saving the files separately. Then, each text document was processed and renamed according to the following structure: fileA_B.C, where:

- A = activity number of the Research and Intervention Supporter in the platform;
- B = identification number of the document, generated by the "LUES Platform;"
- C = file extension.

For instance, a file could be defined as: **file2923_22305.docx**.

The second task was text normalization. The purpose was to (a) standardize the text in lowercase; (b) delete extra spaces between characters; (c) remove any special characters, symbols, duplicate words in a row; and (d) remove accents and numbers. After that, a stop words dictionary was used to remove non-significant aspects of the language components. Since the text reports were written in Brazilian Portuguese, this was the language of the dictionary. Hence, it includes 204 unique words that are considered undesirable for meaningful analysis, such as: "de," "a," "o," "que," "e," "é," "do," "da," "em," "um," and so forth (28). An example of word processing is shown in **Figure 1**.

The third task consisted in removing the duplicate documents. Finally, the fourth one was to make the necessary text corrections, such as grammar error correction and adding missing spaces in some sentences.

2.7. Words and Data Extraction

For capturing words relevance within the text documents, it was necessary to highlight those sentences that came up often and were related to the topic in the texts. Therefore, not attaching any weight to those found in all files–sentences that were obvious in the portrayed context.

To obtain a representation of the topic addressed in the documents and to extract the sentences, the N-gram approach was applied, with $N = 2,3,4$ across the entire group of texts, as well as grouped n-grams using the content analysis method. Finally, the total number of field researchers and reports per region of Brazil was calculated.

2.8. N-grams Extraction: Bigrams, Trigrams, and Quadrigrams

In this step, automatic extractions of keywords formed by a sequence of two words were made. The algorithm also deduced it as automatically relevant in the text documents of Intervention Researchers, the bigrams (N-gram with $N = 2$) (29, 30).

The first step toward extracting the bigrams was to obtain a list of the ordered words from the text (tokens) through the bag-of-words (BoW) approach. Subsequently, the BoW model was converted into a two-word count matrix. Moreover, the Term Frequency—Inverse Document Frequency (TF-IDF) metric was applied (31, 32). Such a technique statistically measures the importance of a word within a text in relation to other pieces within the same database. The value of the word's importance increases proportionally as the total occurrence of that word in the text increases. Thus, the value is compensated by the frequency of that word in the database. That helps discriminate the occurrence of some words that are pretty common but unimportant within the text.

There are two steps to calculating the TF-IDF: (1) separately calculating the TF and then the IDF and (2) multiplying the two parts to find the final value. First, Equation (1) calculates the TF, which measures the probability of a word w occurring in a text *t*.

$$TF_{pt} = \frac{n_{pt}}{\sum_k n_{kt}} \tag{1}$$

Where n_{pt} is the number of times the word *w* occurs in the text *t*; in the denominator, it is the sum of the frequency of all the words in the text. Subsequently, Equation (2) calculates a word's overall relevance across all the texts in the database, IDF.

$$IDF_p = \log \frac{|T|}{\left|t_p : n_{pt} \neq 0\right|} \tag{2}$$

Where $|T|$ indicates the total number of texts in the database and the number of texts in which the word occurs at least a single time. The bigrams were also explored and clustered by region of Brazil. What is more, the same procedure was applied to other extracted n-grams.

When extracting the trigrams (N-gram with N =3), the keywords were formed by a sequence of three words. The same procedure was used for the bigram. However, with the counting matrix being obtained with the formation of three words.

For the extraction of the quadrigrams (N-gram with N=4), the keywords were obtained by a four-word sequence, using the same method as for the bigram, whereby the count matrix was attained by forming four words.

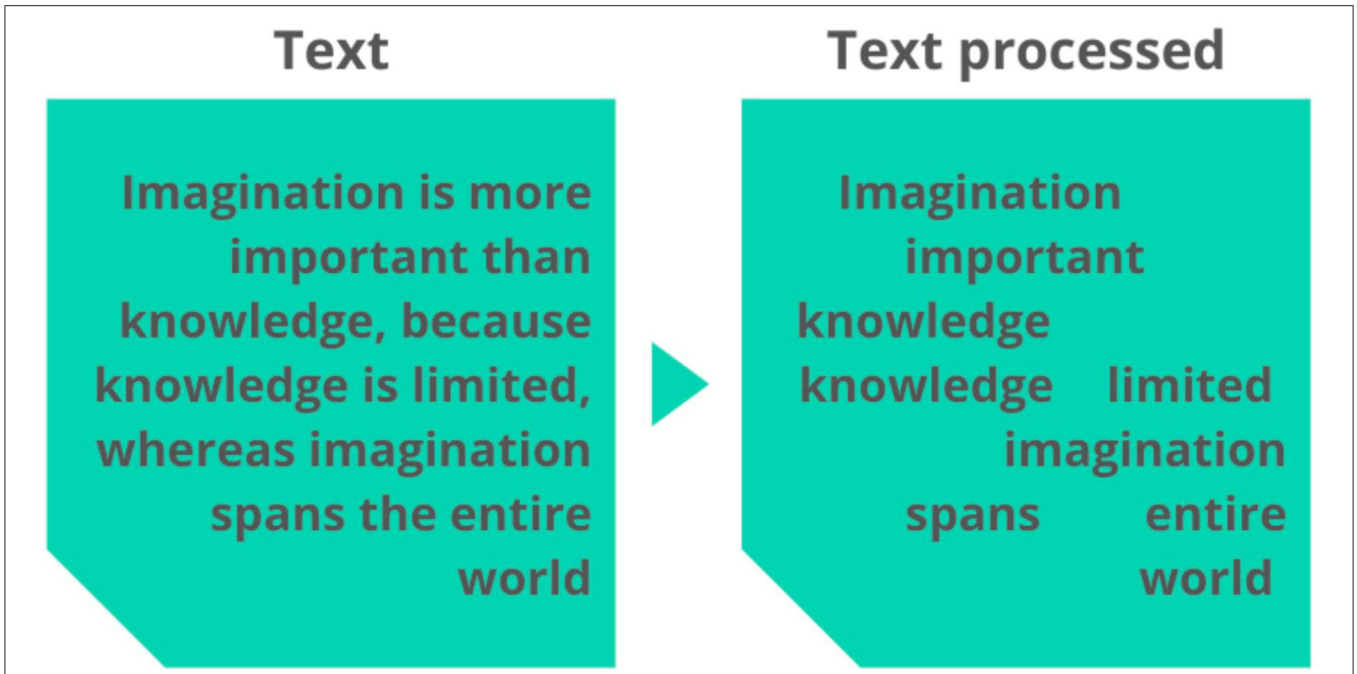


FIGURE 1 | Example of word processing.

N-grams were extracted and grouped by Brazilian region and year. We performed that step based on identifying the structures derived from the levels of health care, syphilis rapid response actions, and other predominant activities observed in reports. Then, the proportion of terms was calculated considering both groups (region and year), as follows:

$$Proportion_{termo} = \frac{\sum N - gram\ per\ report}{number\ of\ reports\ per\ region\ and\ year} \quad (3)$$

2.9. Analysis of Obtained Results: Inference and Interpretation

It is understood that the text production of the field researchers encompasses important content to emphasize the connections of the "Syphilis No" Project in the territory, especially when it comes to CS. In that case, it is about the healthcare field. More specifically, it is about the local response to syphilis and the data, or inputs, on the field researchers' observations and analyses during the project's operationalization. Moreover, in this study, "Syphilis No" is regarded as a public health intervention, based on the public health programs approach by Zulmira Hartz, who analyzes such interventions as actions that favor "adaptive behaviors across distinct human fields or activities" (33).

In the content analysis method, Bardin affirms, deductions are proposed from the significant results and according to the research goals (11). In this research, after the text mining with the formation of the most important bigrams, trigrams, and quadrigrams, a logical sequence was used for the conceptual clustering, depending on the different processes for actions for confronting syphilis that the field researchers have employed. In the conceptual clustering of the terms, the idea was to identify the structures of healthcare levels, programmatic actions in response to syphilis within healthcare networks, and the most prevalent stages identified in the reports produced.

3. RESULTS

Of the 4874 files inserted into the "LUES Platform" in the 2018–2020 period by field researchers, which after pre-processing, totaled 2164, the most are reports, thus totaling 1,019 files, as depicted in **Figure 2**. In addition, other kinds of textual productions were entered into the system, such as images (%), abstracts (%), technical reports (%), communications (%), events or meeting schedules (%), attendance lists (%), and more types of text files (articles, epidemiological bulletins, etc.).

The reports describe the field researchers' intervention activities in the municipalities. So, due to their importance, they were chosen as the basis, or corpus, for applying the content analysis method. It can be seen, from **Figure 3**, that in 2018, 430 reports were produced; in 2019, 355; and in the year 2020, 234 reports.

Figure 4 illustrates the map of Brazil with the geographical distribution of the field researchers. Besides, it describes the total number of reports produced during the three years the field researchers worked on the project and the population size by region. The Southeast of the country is the region with the most field researchers, totaling 22. According to the last census data, it is also where the higher number of reports (629) was created, being the largest region in terms of population size, with 80,364,410 people (34).

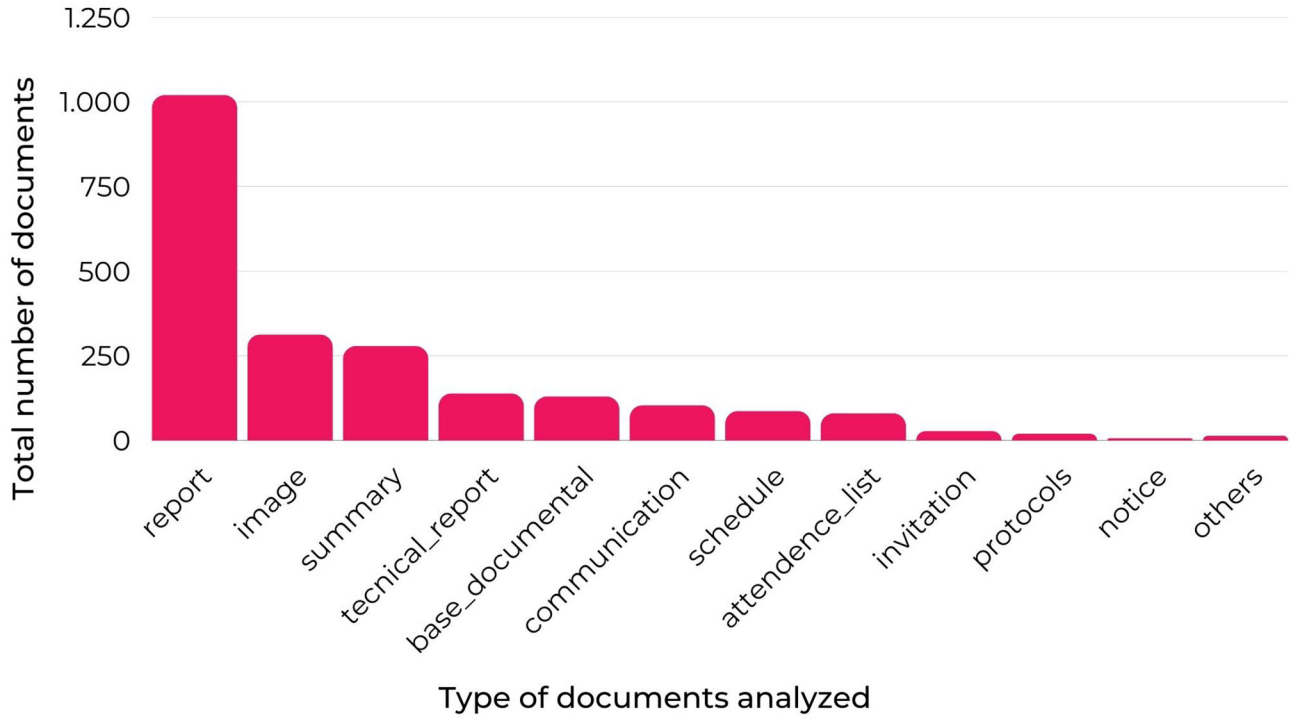


FIGURE 2 | Types of files in the "LUES Platform".

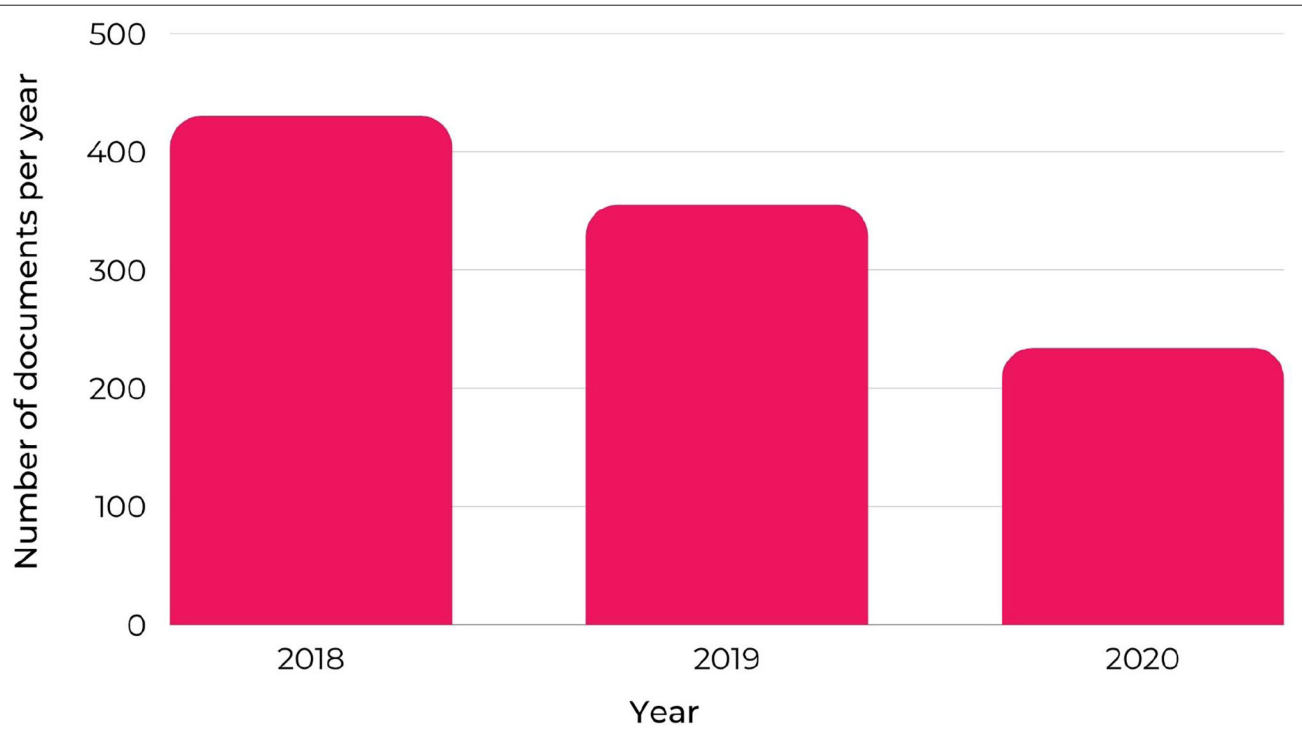


FIGURE 3 | Number of reports produced in Brazil per year.

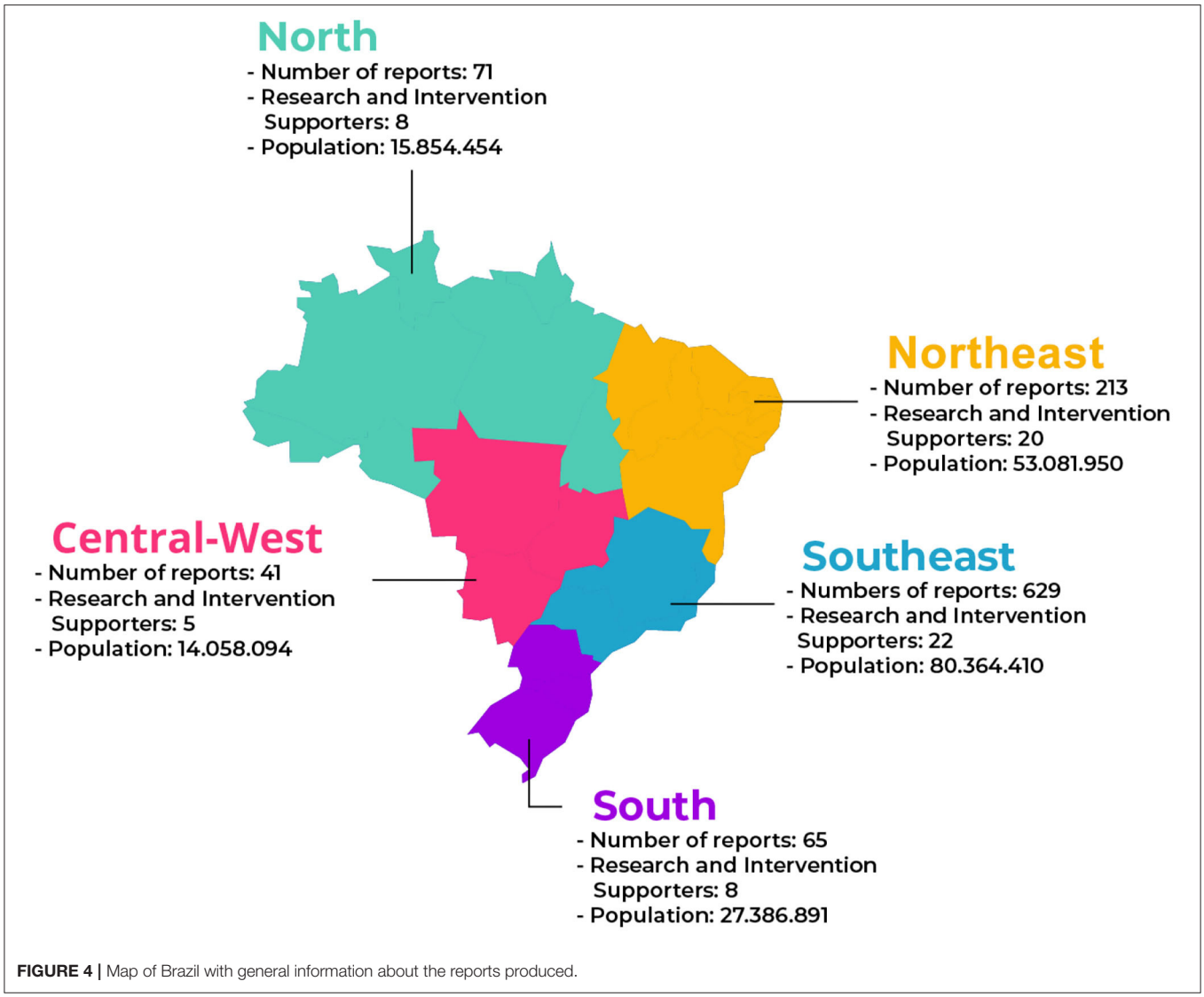


FIGURE 4 | Map of Brazil with general information about the reports produced.

The second region that most produced reports was the Northeast, with a total of 213. It is also the second-largest in relation to field researchers (20) and population size (53,081,950). The South region produced 65 reports, and the Central-West region, 41.

3.1. Analysis of Bigrams

The content analysis based on the bigrams has allowed the verification of which conceptual terms of "Syphilis No" present the most in the reports of the field researchers. The term "congenital syphilis" had the highest frequency. Hence, it signals that this ailment was the most important in the work reports, to the detriment of "syphilis in pregnant women," which had a much lower frequency. Likewise, "acquired syphilis" did not figure among the 20 most present bigrams in the reports (**Figure 5**).

The term "primary care" was the second most frequent. It indicates that this health system component was relevant the most for the supporters when it comes to "confronting syphilis" (third most frequent term). In this vein, "prenatal care" was the most relevant health service, "women's health" was the most significant programmatic area, and "pregnant women" was the most mentioned target audience. Among the other terms, "vertical transmission," "transmission investigation," "syphilis cases" suggest the relevance of epidemiological surveillance efforts in response to CS.

The analysis of the bigrams allowed them to be clustered into exploratory categories related to the response to congenital syphilis and the frequency of these terms in healthcare networks, referring to the original name of the "Syphilis No!" Project. **Figure 6** provides the distribution of the exploratory categories per year and region.

These are considered conceptual terms from the bigrams: "Confronting syphilis in prenatal care in primary health care" (**Figure 6A**); "investigation of transmission of congenital syphilis in the territory" (**Figure 6B**); "epidemiological surveillance within the municipal health plan" (**Figure 6C**); "syphilis in

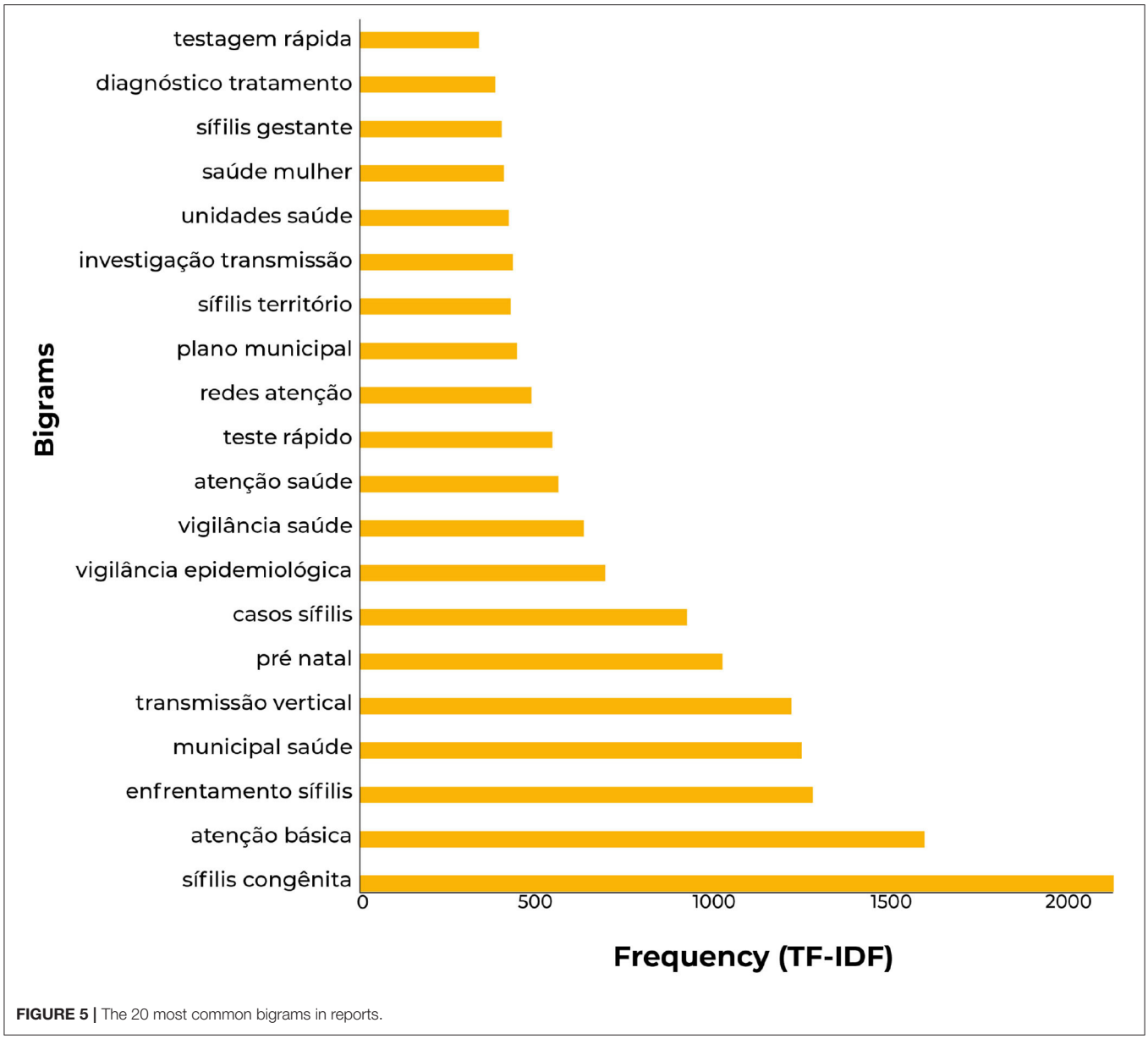


FIGURE 5 | The 20 most common bigrams in reports.

pregnant women in women's health care" (**Figure 6D**); "rapid testing, diagnosis, and treatment" (**Figure 6E**).

"Confronting syphilis in prenatal care in primary health care" (**Figure 6A**) shows that the top activities occurred in 2018 in most regions, except for the South and Southeast regions. For the "investigation of transmission of congenital syphilis in the territory" (**Figure 6B**), the Central-West and North regions indicated high proportions at the very beginning of the project. As for the "epidemiological surveillance within the municipal health plan" (**Figure 6C**), its best distribution was in 2018 in the Central-West and Southeast regions. For the "syphilis in pregnant women in women's health care" (**Figure 6D**), the South region exhibited the highest activity in 2019. Finally, the highest proportion of "rapid testing, diagnosis and treatment" (**Figure 6E**) was in 2019 in all regions, with the South region being prominent.

3.2. Analysis of Trigrams

The 20 most relevant trigrams extracted in the analysis have been pointed out in **Figure 7**, with greater relevance to "investigation vertical transmission," "congenital syphilis cases," "vertical transmission syphilis," and "syphilis healthcare networks." Such trigrams reveal that the response to MTCT (vertical transmission) of congenital syphilis–including the investigation of transmission within healthcare networks–was considered a priority in the reports.

These are considered conceptual terms from the trigrams: "Committee for investigation of MTCT of congenital syphilis"

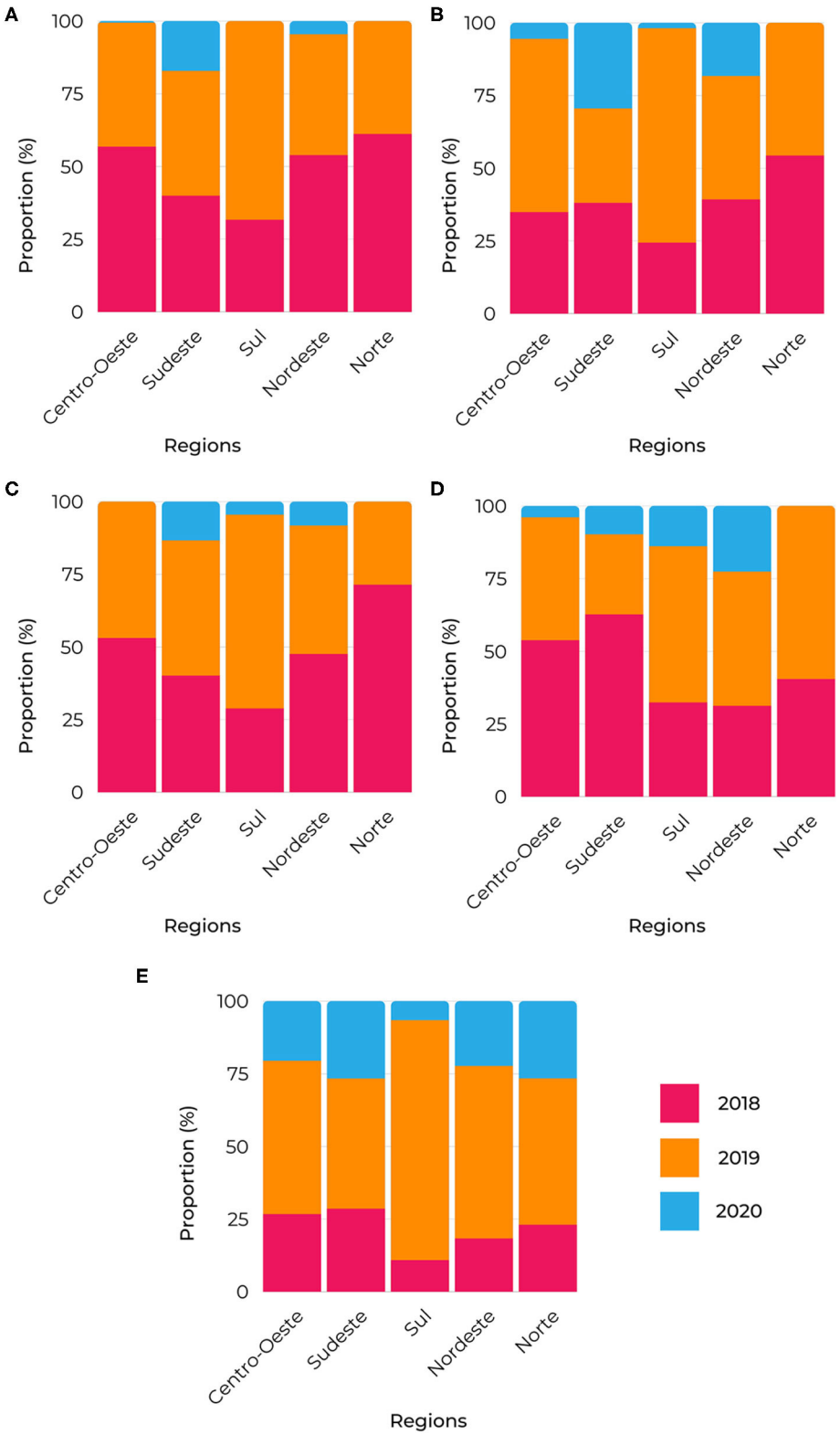


FIGURE 6 | Bigrams proportion of relevant terms segmented by region. **(A)** primary care, Confronting syphilis, syphilis cases and prenatal care; **(B)** care networks, health units, women's health and pregnant syphilis; **(C)** investigation transmission, syphilis territory, and congenital syphilis; **(D)** municipal plan, epidemiological surveillance, and health surveillance; **(E)** rapid testing and diagnosis treatment.

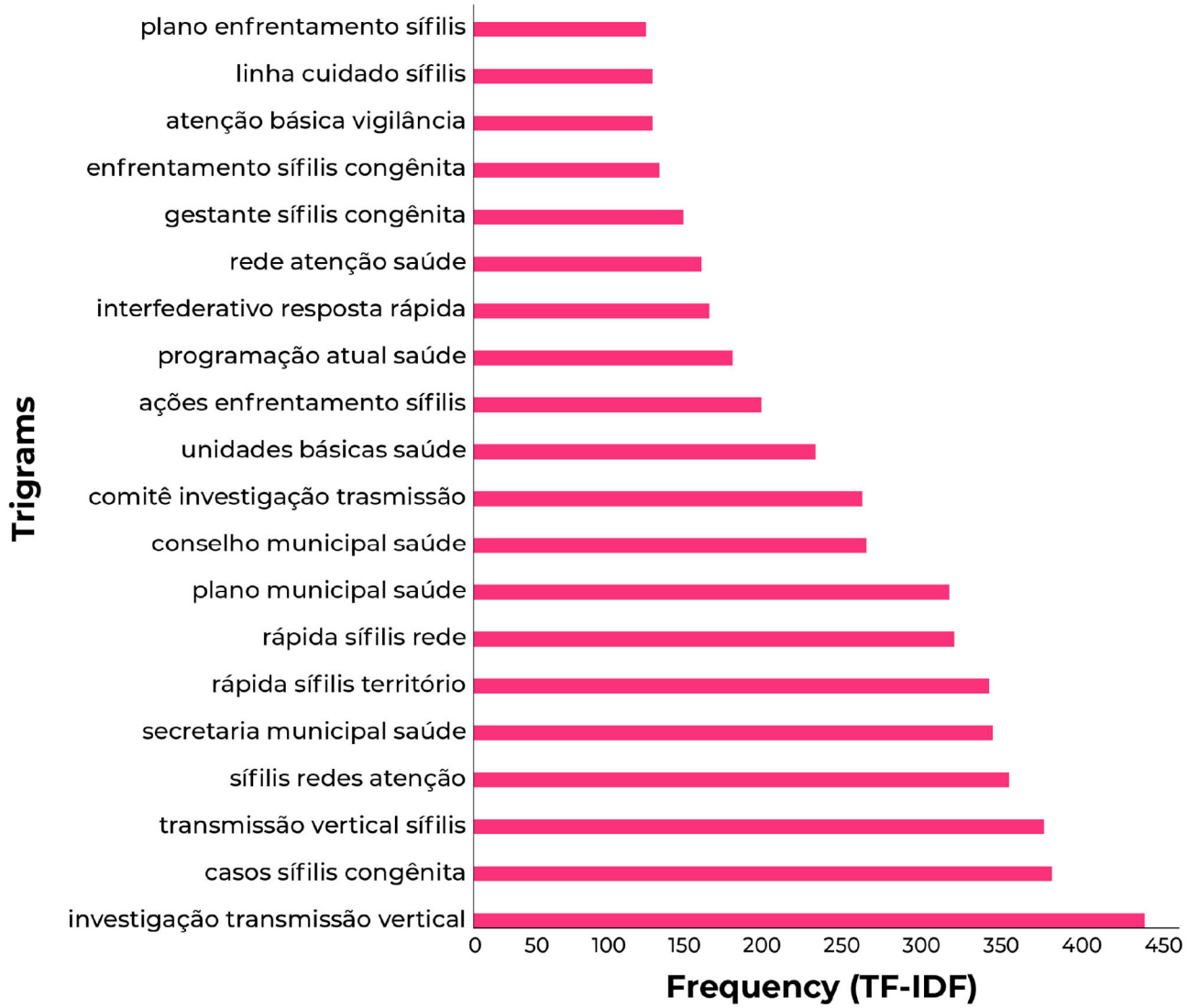


FIGURE 7 | The 20 most common trigrams in reports.

(**Figure 8A**), "congenital syphilis in municipalities' planning and yearly health programming" (**Figure 8B**), and "surveillance and attention in the lines of care for congenital syphilis and syphilis in pregnant women in primary centers of healthcare networks" (**Figure 8C**).

The analysis based on the trigram clusters corroborates the tendency of the work of the supporters within the scope of the "investigation of vertical transmission" of "congenital syphilis" (**Figures 8A–C**). Further, "pregnant women" remains as the most crucial target population (**Figure 8C**), and the "municipal health plan" emerges as a relevant instrument, adding to it the "yearly health programming" and the "municipal health councils."

Also, in **Figure 8**, the most significant activities for the "Committee for investigation of vertical transmission of congenital syphilis" (**Figure 8A**) and for "congenital syphilis in municipalities' planning and yearly health programming" (**Figure 8B**) occurred in 2018 for virtually all regions. Lastly, the topmost activity of the conceptual term "surveillance and attention in the lines of care for congenital syphilis and syphilis in pregnant women in primary centers of healthcare networks" (**Figure 8C**) was in 2019 across all the regions.

3.3. Analysis of Quadrigrams

The extracted quadrigrams can be seen in **Figure 9**. The highest frequency is for part of the full name of the "Syphilis No" project ("rapid response syphilis project").

The remaining quadrigrams confirm the field researchers' priority for investigating vertical transmission and the committee as a substantial means or service for response to CS in the territory. Besides, four conceptual terms were generated and distributed according to year and region (see **Figure 10**).

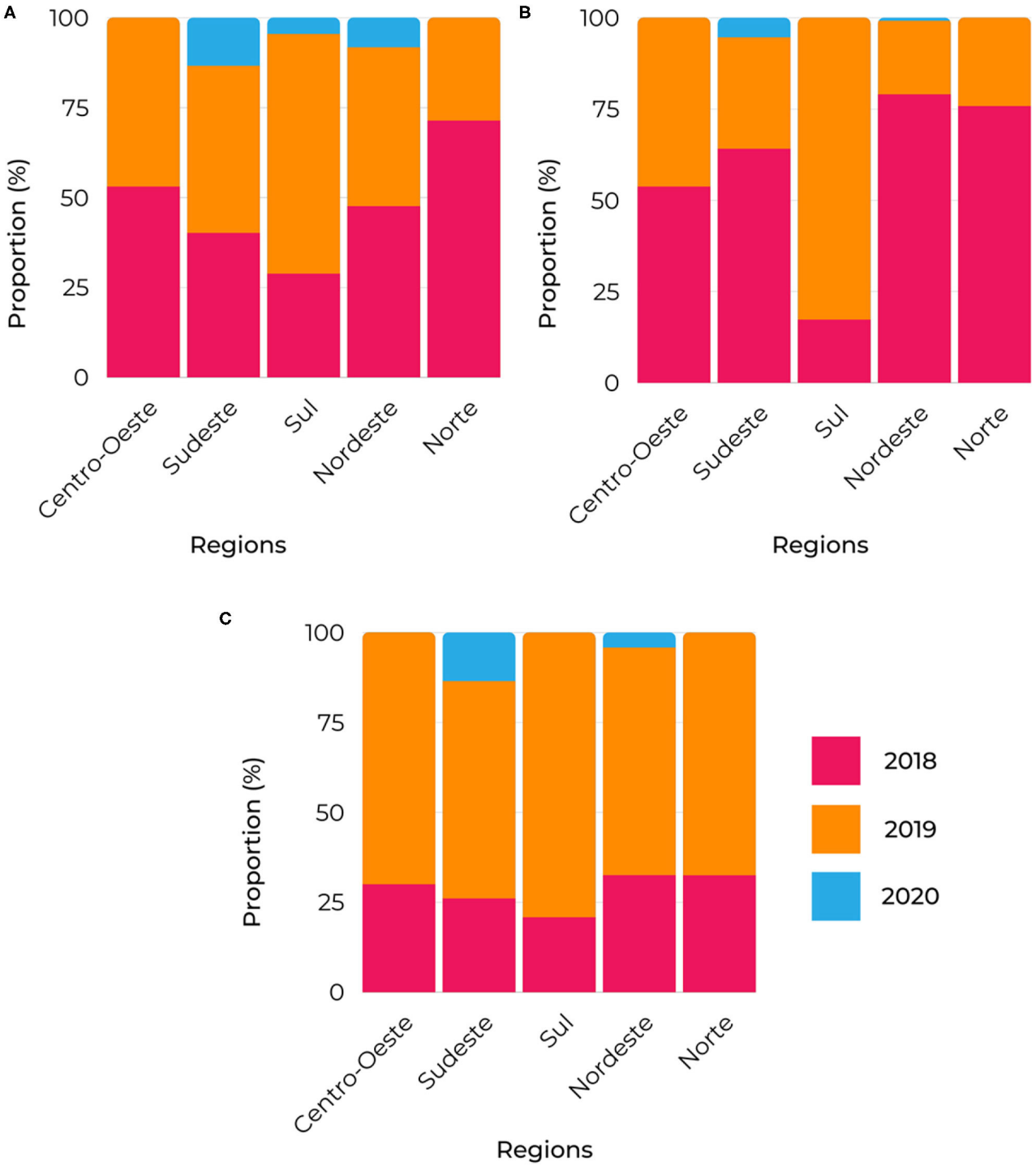


FIGURE 8 | Trigrams proportion of relevant terms segmented by region. **(A)** investigation vertical transmission, congenital syphilis cases, vertical transmission syphilis and investigation committee transmission; **(B)** municipal health plan, municipal health secretariat, municipal health council, annual health program and confrontation congenital syphilis; **(C)** primary health center, rapid syphilis network, rapid syphilis territory, congenital syphilis pregnant, health care network, primary care surveillance and line care syphilis.

In **Figure 10A**, the concept of "implementation of the rapid response project through strategic actions of the research and intervention supporter" has prevailed. As a result, this term attained its most outstanding distribution in all regions in 2019. The second conceptual term spawned was "rapid response to syphilis leads to advancements in healthcare networks regarding

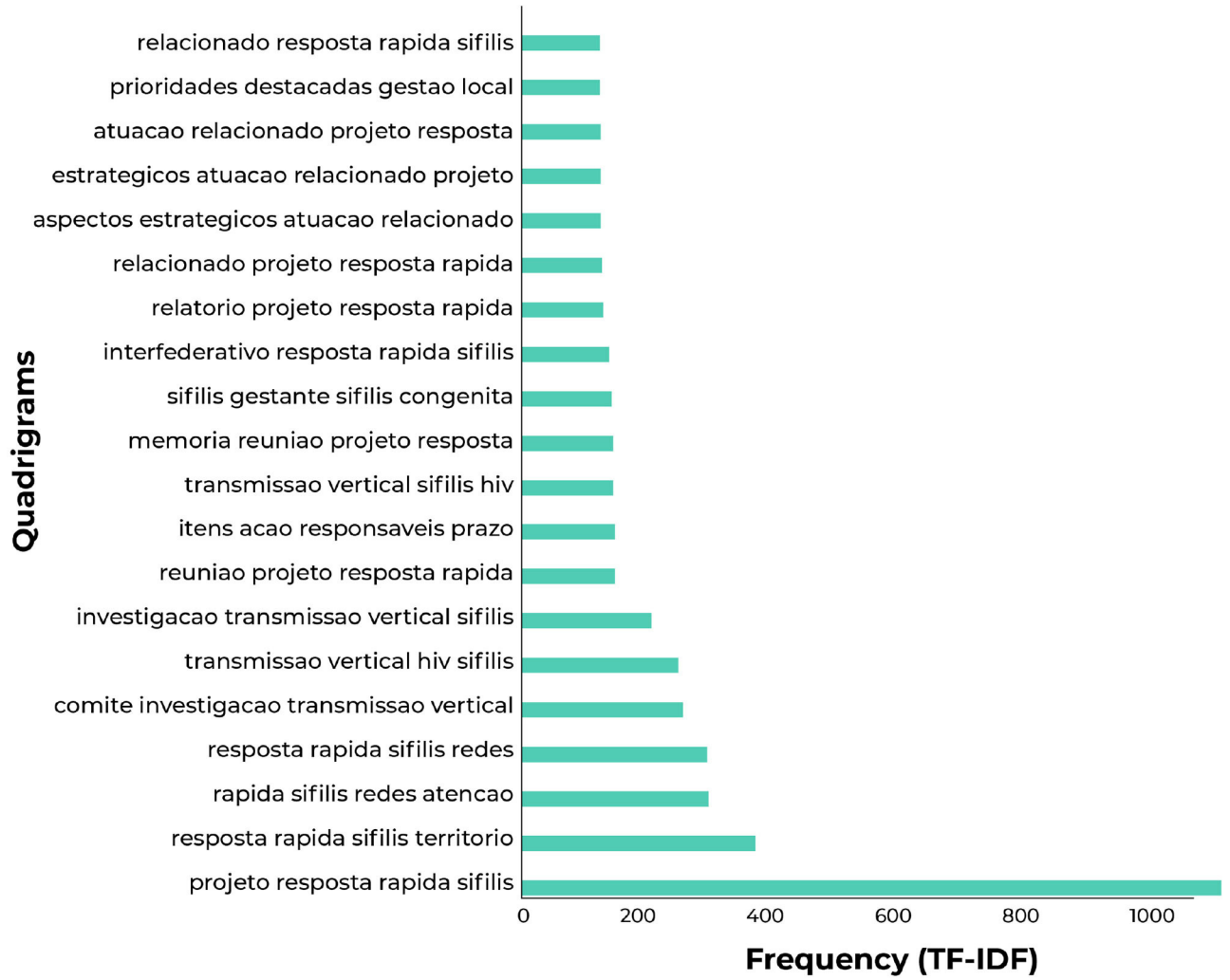


FIGURE 9 | The 20 most common quadrigrams in reports.

congenital syphilis," with predominant distribution in almost all regions at the project's onset (**Figure 10B**).

In **Figure 10C**, the conceptual term was "local management priorities for combating syphilis in pregnant women and congenital syphilis," with its almost total distribution in 2019 in virtually all regions. Last, the conceptual term "committee for the investigation of vertical transmission led by syphilis rapid response" had its highest distribution in almost all regions at the beginning of the project (**Figure 10D**).

It was noted that the quadrigrams have further underscored the possibilities for correlations between the activities of field researchers related to CS and the dimensions of Governance and Management, Surveillance and Comprehensive Care, of the SNP.

Figure 11 presents the interrelatedness between the conceptual terms formed in light of the objectives of the "Syphilis No" Project for the bigrams, trigrams, and quadrigrams. Still, the interpretation of these establishes at least five analytical categories that are considered representative of the associations between the research supporter's intervention work and the goals of the project:

1. Confronting syphilis during prenatal care in primary health care;
2. Committee for investigation of congenital syphilis in the territory;
3. Municipal plan for monitoring and investigating syphilis cases through health surveillance;
4. Women's health care networks for addressing syphilis in pregnant;
5. Diagnosis and treatment with a focus on rapid testing.

Thus, it can be inferred that the work of SNP field researchers was primarily that of intervention for confronting the response to congenital syphilis in the territory where they worked, following the line for eradication of CS by each municipality.

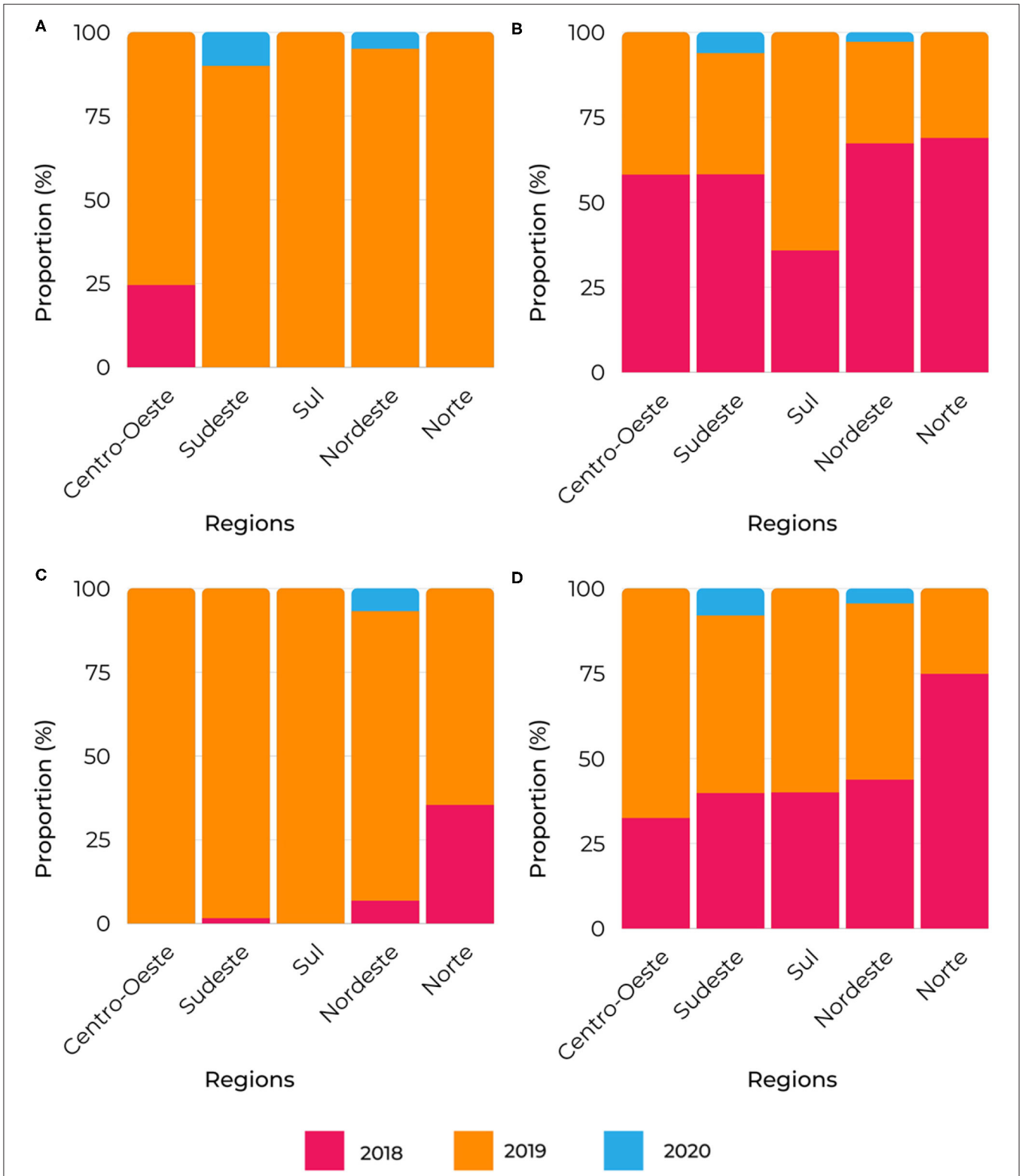


FIGURE 10 | Quadrigrams proportion of relevant terms segmented by region. **(A)** rapid response project implementation, supporter related response project, main activities developed territory, strategic aspects action supporter; **(B)** rapid syphilis care networks, congenital syphilis lower year, advances rapid response syphilis, related rapid response project; **(C)** highlighted priorities local management, highlighted local management related, syphilis pregnant congenital syphilis, national day against syphilis; **(D)** investigation vertical transmission committee, investigation vertical syphilis transmission, related rapid response syphilis.

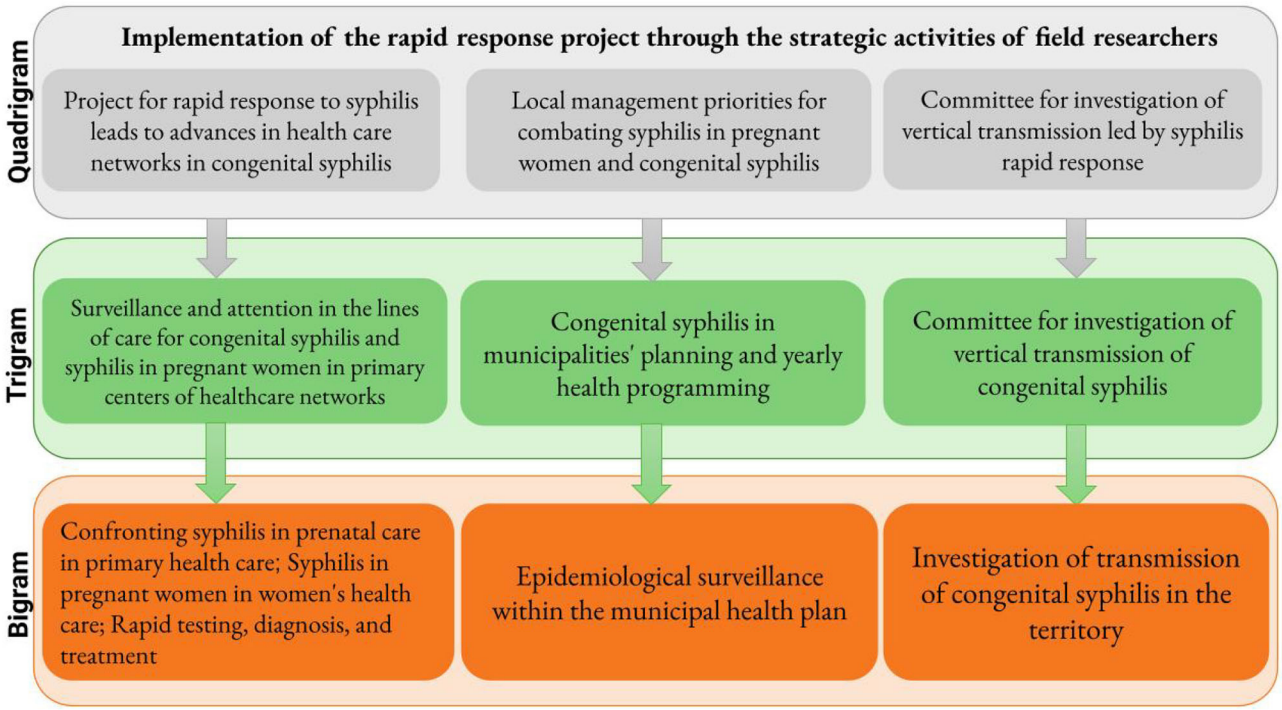


FIGURE 11 | Flow diagram of literature search.

4. DISCUSSION

The research has evidenced that using the text mining technique has enabled the identification of patterns related to the action or intervention of the field researchers' work regarding syphilis in the priority municipalities. Likewise, it has facilitated the measurement and comparison of the researchers' efforts by region of the country. Five analytical categories were determined. These demonstrate significant and representative associations between the articulation work of the Research and Intervention Supporter and the objectives of the "Syphilis No!" Project. Specifically, the categories were:

- Confronting syphilis during prenatal care in primary health care;
- Investigation committee for congenital syphilis in the territory;
- Municipal plan for monitoring and investigating syphilis cases through health surveillance;
- Women's healthcare networks for syphilis in pregnant;
- Diagnosis and treatment with a focus on rapid testing.

It is worth remarking that the field researchers' performance in the project has emerged to intervene and help the technical teams and local managers. So, one of the main actions was to reduce the high rates of congenital syphilis in priority municipalities (35). For instance, in Brazil, from 2010 to 2017, the rate of CS was from 2.0 cases per 1,000 live births to 8.8 per 1,000 live births, respectively. Thus, an increase of 338%. There was a significant upward trend in congenital syphilis in the same interval, with an AAPC of 15.75%, with no prospect of a change in the pattern (8).

In 2018, CS rates hit an all-time high, reaching the staggering incidence of 9 cases per 1,000 live births. Nevertheless, according to the most recent epidemiological bulletin, after the beginning of the SNP, the congenital syphilis rates have decreased to 8.2 cases per 1000 live births (36). As a result, there was a change in the growth trend. It went from 15.75% to a reduction of -8.8% in the annual incidence rate of CS, with a prospect for reduction over the next few years. It may reach the 90% reduction target established by PAHO in almost four years (37). In a short span, the nationwide decrease in congenital syphilis rates can be credited to the "Syphilis No" intervention model. In this manner, the epidemiological impact proved to be effective, especially from identifying the thematic areas of the intervention included in the reports of the RIS.

The research done by Pinto and colleagues assessed the impact of actions developed before and after the SNP in Brazilian municipalities based on interrupted time series analysis. These scholars have concluded that the intervention project model contributed significantly to CS rates decrease in priority areas. The conceptual terms built from the researchers' reports can point to the most significant actions that resulted in the reduced CS rates in the intervention territories.

Our findings made it possible to identify a tendency to terms clustering directed to the programmatic actions related to congenital syphilis in the priority municipalities. The trigrams and quadrigrams have solidified this trend observed in the

bigrams. In addition, they went beyond with information about the work of the supporters per axis of the project.

The Pan American Health Organization's (PAHO) Framework for Elimination of Mother-to-Child Transmission of STIs in the Americas has as one of the conceptual framework's dimensions to integrate policies targeted at women and children with the related health services. That includes the prevention stages that precede pregnancy and the health of women and children after perinatal care and childbirth (37). The results of this article have evidenced that the integration between surveillance and care is part of the field researchers' work. It is corroborated by the presence of the investigation committees of MTCT. Thus, showing that the response to syphilis in Brazil draws on a strategy not foreseen in PAHO's EMTCT Plus, and therefore is innovative. Furthermore, the data reveal that the principal intervention of the supporters was in the committees' efforts. It also indicates the strategic importance of such surveillance and comprehensive care policy to eliminate CS.

The Brazilian health system has its specific protocols for investigating and intervening in identified cases of CS. This work is conducted by the investigation committees of vertical transmission or multi-professional teams of specialists. These were formed to map the problems related to mother-to-child transmission and propose solutions based on a pre-established investigation protocol to mitigate CS. The committees are integrated into SUS as spaces for technical, confidential, non-coercive, and non-punitive activities and may be linked to municipal or state health management (38, 39).

Research findings also demonstrated that these research committees played a substantial role in the "Syphilis No!" Project, cited mainly in 2018, with prominence in the Northern region. In addition, the investigation of MTCT, shown in the bigram, had its different distribution, registering the highest proportions for the Northeast and Southeast regions, keeping these activities practically constant in the 2018-2020 time frame.

Such regional discrepancies can be attributed to the organizational complexity of Brazilian healthcare networks (40), whose implementation of committees for investigation of vertical transmission will depend on each local government in a decentralized approach. Therefore, the regions with weaker consolidation of the committees have probably required more in-depth attention from the field researchers at the beginning of the project, as observed in the Northern region (41). This fact may have influenced the effectiveness of the efforts to tackle CS in that region.

The process of mining the reports of field researchers' intervention activities has also highlighted the actions to confront syphilis in prenatal care, primary health care, and syphilis in pregnant women in women's health care. In Brazil, for example, among the experiences of the Ministry of Health considered specific to the prevention of mother-to-child transmission (PMTCT) that preceded the SNP' were the project "Nascer Maternidades" (2002) and the Operational Plan for Reducing HIV and Syphilis (2007).

The project "Nascer Maternidades" aim was to ensure the identification of the pregnant woman's serological status for HIV infection during labor (42). Moreover, the Operational Plan for the Reduction of HIV and Syphilis aimed at expanding the coverage of HIV and syphilis testing in prenatal care and guaranteeing actions agreed upon with the states and municipalities since the 1990s by the National HIV/Aids Program. Such actions were strongly related to improving HIV/AIDS indicators, having as primary strategies the acquisition and distribution of antiretroviral drug supplies; infant formula for breastfeeding HIV-exposed children; and determining the routine management of HIV during pregnancy, labor and delivery, and postpartum (43).

These policies were induced through purchasing the principal supplies and the accountability for their use by the three federative entities of SUS. Therefore, it can be concluded that these strategies, on the one hand, have contributed to PMTCT of HIV. However, on the other hand, they have not been able to curb the progression of the syphilis epidemic in Brazil, according to epidemiological indicators before the project's start (36).

Hence, the SNP is probably more advanced than previous projects because it introduced a proposal to induce public health policies in an integrated and cooperative way between surveillance and care as to syphilis in the territory—having RIS a pivotal role as articulating agents. Furthermore, the data suggest that these supporters have managed to align the activities and interventions of projects related to women's health that already existed in the intervention sites with the objectives of the "Syphilis No!" Project. Specifically, this research found that the most relevant interventions were observed for the target population of pregnant women in healthcare networks.

Regarding the distribution of the supporters' interventions by region, it stands out that prevention and assistance efforts were least reported in the South region at the beginning of the project. However, that distribution increased in 2019. It is worth pointing out that the same region has the highest growth trends of CS in Brazil (8). This fact may have contributed to more resistance from local managers at the beginning of the project's implementation. So, it can be inferred that the field researchers from the South have experienced more difficulty, when compared to the other regions, in their insertion in the territory.

Rapid testing for syphilis was another aspect of health care highlighted in the field researchers' production. For instance, a study developed by Santos and collaborators (44) to identify factors that influenced the fight against syphilis emphasized that rapid testing plays a crucial role in fighting syphilis in primary health care. Another study, developed by Roncalli and colleagues, showed that, from 2018 to 2019, there was the highest availability of rapid testing to confront congenital syphilis in history, independent of other confounding factors.

In this way, it can be surmised that the intervention activities of the field researchers contributed to the strengthening of Point-of-care rapid testing (45, 46) measures for CS prevention, which remained nearly constant throughout the production of the reports in most regions in the 2018-2020 period.

The intervention activities of the field researchers of "Syphilis No!" Project may have also contributed to the decrease in the number of congenital syphilis hospitalizations nationwide. Besides, hospitalizations rates for treatment of CS were verified to have significantly dropped from May 2018 to December 2019 (10). Such data corroborate this study's results, whose assistance and surveillance actions were featured

in the field reports' production and the formation of bigrams and trigrams.

Some limitations can be identified in this study. One example is that the formation of bigrams, trigrams, and quadrigrams was mined from an algorithm that highlighted the most substantial activities of the field researchers. Consequently, that may rule out activities that have been developed with the local health services management and had some significant impact on the changes in the work processes for syphilis prevention, such as words related to the education and communication axis of the project. Nevertheless, despite such a limitation, this research distinguished which efforts were most important in contributing to the current changes in the epidemiological indicators of syphilis in Brazil.

Although field researchers provided different types of data, e.g., images and videos, these were not explored as this study focused on the data mining method for written texts. As a result, such data were removed since they were out of scope and would require computer vision and digital image processing methods for information extraction. Future research could explore TDM applied to alternative types of data.

5. CONCLUSION

The study has found that when combined with the conventional content analysis method, text mining can address public health research subjects that comprise large volumes of data. This computational method made it possible to extract, from the SNP, the intervention activities of field researchers. Furthermore, it subsidized inferences on how the project's strategies may have led to a drop in CS cases in the priority municipalities.

This work offers contributions permeating several fields: (i) analysis of large sets of data through TDM techniques; (ii) strengthening the prospects for using TDM in conjunction with content analysis, broadly applied in health research involving text production to support the assessment of interventionist actions by field researchers in large-scale national projects; and (iii) evidence that efforts toward articulating and promoting public health policies in the territory have had positive effects in managing the response to congenital syphilis. In addition, the text mining of the reports pointed out there is an articulation among the actions of the field researchers regarding CS in three out of the four dimensions of the project aixes which are governance and management; surveillance and comprehensive care.

Another highlight was that the work of the field researchers revealed the links between the project's objectives and the confrontation of syphilis in the territory. Hence, it demonstrated through analytical categories how the field researchers' priority action was developed and their contribution to reducing indicators of CS in Brazil. Therefore, it is possible to infer that the field researchers' activities, highlighted in the reports, may have induced public health policies in the territory by supporting prevention and health promotion efforts to combat CS as a priority by local health services managers. Hopefully, further work will determine which field researchers' actions have positively influenced other essential indicators in the fight against syphilis. These steps are detailed in the following sessions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: http://vigilanciasaude.ufrn.br/files/anonymous_reports.csv.

AUTHOR CONTRIBUTIONS

MR: data processing, article writing and research methodology development. RF: data processing and research methodology development. MS and TL: article writing and research methodology development. ASPM: literature review and article writing. AC-O, AEM, CO, HO, and CG: guidelines and review of the article. RP: writing and reviewing the article. DB: literature review, writing and reviewing the article. RV: article writing, review, research coordination, and research methodology development. All authors read and approved the final manuscript.

FUNDING

This work was supported by Brazilian Ministry of Health–Number's project: 54/2017.

ACKNOWLEDGMENTS

We kindly thanks the Laboratory of Technological Innovation in Health of the Federal University of Rio Grande do Norte by supporting the research.

REFERENCES

1. Crosby AW Jr. The early history of syphilis: a reappraisal. *Am Anthropol.* (1969) 71:218–27.
2. Rothschild BM. History of syphilis. *Clin Infect. Dis.* (2005) 40:1454–63. doi: 10.1086/429626
3. Kojima N, Klausner JD. An update on the global epidemiology of syphilis. *Curr Epidemiol Rep.* (2018) 5:24–38. doi: 10.1007/s40471-018-0138-z
4. Korenromp EL, Rowley J, Alonso M, Mello MB, Wijesooriya NS, Mahiané SG, et al. Global burden of maternal and congenital syphilis and associated adverse birth outcomes–Estimates for 2016 and progress since 2012. *PLoS ONE.* (2019) 14:e0211720. doi: 10.1371/journal.pone.0211720
5. Forrai J. History of different therapeutics of venereal disease before the discovery of penicillin. In: *Syphilis-Recognition, Description and Diagnosis.* London: In Tech (2011). p. 37–58.
6. Sarbu I, Matei C, Benea V, Georgescu S. Brief history of syphilis. *J Med Life.* (2014) 7, 4.
7. Nurse-Findlay S, Taylor MM, Savage M, Mello MB, Saliyou S, Lavayen M, et al. Shortages of benzathine penicillin for prevention of mother-to-child transmission of syphilis: an evaluation from multi-country

surveys and stakeholder interviews. *PLoS Med.* (2017) 14:e1002473. doi: 10.1371/journal.pmed.1002473
8. Marques dos Santos M, Lopes AKB, Roncalli AG, Lima KCd. Trends of syphilis in Brazil: a growth portrait of the treponemic epidemic. *PLoS ONE.* (2020) 15:e0231029. doi: 10.1371/journal.pone.0231029
9. Pinto R, Valentim R, da Silva LF, de Souza GF, Lima TGFdMS, de Oliveira CAP, et al. Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in Brazil. *Lancet Regional Health.* (2021) 7:100163. doi: 10.1016/j.lana.2021.100163
10. de Andrade IGM, de Medeiros Valentim RA, de Oliveira CAP. The influence of the No Syphilis Project on congenital syphilis admissions between 2018 and 2019. DST j bras doenças sex transm. 2020;p. 1–6.
11. Bardin L. *Content Analysis.* . São Paulo, CA: Edições. (2011).
12. Souza MARd, Wall ML, Thuler ACdMC, Lowen IMV, Peres AM. The use of IRAMUTEQ software for data analysis in qualitative research. *Revista da Escola de Enfermagem da USP.* (2018) 52:e03353. doi: 10.1590/S1980-220X2017015003353
13. Amaral-Rosa MP. Considerations on the use of IRAMUTEQ software for qualitative data analysis. *Revista da Escola de Enfermagem da USP.* (2019) 53:e03468. doi: 10.1590/S1980-220X2019ce0103468
14. Kumar S, Kar AK, Ilavarasan PV. Applications of text mining in services management: a systematic literature review. *Int J Inf Manag Data Insights.* (2021) 1:100008. doi: 10.1016/j.jjimei.2021.100008
15. Kar AK, Dwivedi YK. Theory building with big data-driven research–Moving away from the "What" towards the "Why". *Int J Inf Manag.* (2020) 54:102205. doi: 10.1016/j.ijinfomgt.2020.102205
16. Li Y, Rapkin B, Atkinson TM, Schofield E, Bochner BH. Leveraging latent dirichlet allocation in processing free-text personal goals among patients undergoing bladder cancer surgery. *Qual Life Res.* (2019) 28, 1441–55. doi: 10.1007/s11136-019-02132-w
17. Lee H, Shimotakahara R, Fukada A, Shinbashi S, Ogata S. Impact of differences in clinical training methods on generic skills development of nursing students: a text mining analysis study. *Heliyon.* (2019) 5:e01285. doi: 10.1016/j.heliyon.2019.e01285
18. El Kah A, Zeroual I. A review on applied natural language processing to electronic health records. In: *2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA).* Sana'a: IEEE (2021). p. 1–6.
19. Valentim RAM, Caldeira-Silva GJP, da Silva RD, Albuquerque GA, de Andrade IGM, Sales-Moioli AIL, et al. Stochastic Petri net model describing the relationship between reported maternal and congenital syphilis cases in Brazil. *BMC Med Inform Decis Mak.* (2022) 22:40. doi: 10.1186/s12911-022-01773-1
20. Demner-Fushman D, Elhadad N, Friedman C. Natural language processing for health-related texts. In: *Biomedical Informatics.* New York, NY: Springer (2021). p. 241–72.
21. Rocha MAd, Nóbrega GÂSd, de Medeiros Valentim RA, Alves LPC. A text as unique as fingerprint: AVASUS text analysis and authorship recognition. In: *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems.* New York, NY (2020). p. 1–8.
22. Stewart R, Velupillai S. Applied natural language processing in mental health big data. *Neuropsychopharmacology.* (2021) 46:252. doi: 10.1038/s41386-020-00842-1
23. Javed AR, Sarwar MU, Beg MO, Asim M, Baker T, Tawfik H. A collaborative healthcare framework for shared healthcare plan with ambient intelligence. *Hum Centric Comput Inf Sci.* (2020) 10:1–21. doi: 10.1186/s13673-020-00245-7
24. Vespestad MK, Clancy A. Exploring the use of content analysis methodology in consumer research. *J Retailing Consum Services.* (2020) 59:102427. doi: 10.1016/j.jretconser.2020.102427
25. de Morais PSG, Silva RDd, Filho JAP, de Medeiros Valentim RA, Coutinho KD, de Oliveira CAP, et al. Strategies for content recommendation in the Brazilian rapid response to syphilis project. In: *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems.* New York, NY (2020). p. 1–6.
26. Priamo V, dos Santos SC, dos Santos JS. PISTAS PARA O TRABALHO DO APOIO NO PROJETO "SÍFILIS NÃO". In: *Revista Brasileira de Inovação Tecnológica em Saúde-ISSN: 2236-1103.* (2020) 10:11.
27. *Shafranovich Y. Common Format and MIME Type for Comma-Separated Values (CSV) Files.* RFC 4180, October, (2005).
28. Uysal AK, Gunal S. The impact of preprocessing on text classification. *Inf Process Manag.* (2014) 50:104–12. doi: 10.1016/j.ipm.2013.08.006
29. SUHASINI V DNV. A Hybrid TF-IDF and N-Grams Based Feature Extraction Approach for Accurate Detection of Fake News on Twitter Data. *Turkish J Comput Mathe Educ.* (2021) 12:5710–23.
30. Rajput S, Gahoi A, Reddy M, Sharma DM. N-Grams TextRank: a novel domain keyword extraction technique. In: *17th International Conference on Natural Language Processing.* Patna (2020). p. 9.
31. Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. *Inf Process Manag.* (1988) 24:513–23.
32. Thakkar A, Chaudhari K. Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks. *Appl Soft Comput.* (2020) 96:106684. doi: 10.1016/j.asoc.2020.106684
33. Hartz ZMdA. Avaliação dos programas de saúde: perspectivas teórico metodológicas e políticas institucionais. *Ciência saúde coletiva.* (1999) 4:341–53.
34. Censo I. "Instituto Brasileiro de Geografia e Estatística. 2010," *Sinopse por Setores Disponível.* Rio de Janeiro: IBGE. Available online at: https://censo2010ibgegovbr/sinopseporsetores. (2000).
35. BRASIL MdS. Seminário apresenta projeto "Resposta Rápida à Sífilis nas Redes de Atenção" a profissionais de saúde. (2018) p. 1. Available online at: http://www.aids.gov.br/pt-br/noticias/seminario-apresenta-projeto-resposta-rapida-sifilis-nas-redes-de-atencao-profissionais-de (accessed December 28, 2021).
36. da S BM. Boletim Epidemiológico de Sífilis. Secr Vigilância em Saúde [Internet]. (2020).
37. de S OOA. Politica para um enfoque Integrado e sustentável Visando as doenças Transmissiveis nas Americas [Internet]. (2019).
38. da Saúde (BR) M. Protocolo de infestigação de transmissão vertical. Ministério da Saúde Brasília. (2014).
39. *da Saúde (BR) Secretaria de Vigilância em Saúde Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis M. Protocolo clínico e diretrizes terapêuticas para atenção integral às pessoas com infecções sexualmente transmissíveis (IST).* Brasília: Ministério da Saúde Brasília. (2019).
40. Oliveira APCd, Gabriel M, Poz MRD, Dussault G. Challenges for ensuring availability and accessibility to health care services under Brazil's Unified Health System (SUS). *Ciencia Saude Coletiva.* (2017) 22:1165–80. doi: 10.1590/1413-81232017224.31382016
41. Garnelo L, Sousa ABL, Silva CdOd. Health regionalization in Amazonas: progress and challenges. *Ciencia Saude Coletiva.* (2017) 22:1225–34. doi: 10.1590/1413-81232017224.27082016
42. Santos EMd, Reis AC, Westman S, Alves RG, et al. *Avaliação do grau de implantação do programa de controle da transmissão vertical do HIV em maternidades do" Projeto Nascer."* Brasília: Ministério da Saúde. (2010).
43. da Saúde Secretaria de Vigilância em Saúde BM. Programa Nacional de DST/Aids. Plano Operacional. Redução da Transmissão Vertical do HIV e da Sífilis. (2007). Available online at: https://bvsms.saude.gov.br/bvs/publicacoes/plano_operacional_WEB.pdf (accessed December 28, 2021).
44. Santos MMd, Rosendo TMSdS, Lopes AKB, Roncalli AG, Lima KCd. Weaknesses in primary health care favor the growth of acquired syphilis. *PLoS Negl Trop Dis.* (2021) 15:e0009085. doi: 10.1371/journal.pntd.0009085
45. Organization WH. World Health Organization. Point-Of-Care Diagnostic Tests (POCTs) for Sexually Transmitted Infections (STIs). 2021. (2021). Available online at: https://www.who.int/reproductivehealth/topics/rtis/pocts/en/ (accessed December 28, 2021).
46. Angel-Müller E, Grillo-Ardila CF, Amaya-Guio J, Torres-Montañez NA, Vasquez-Velez LF. Point of care rapid test for diagnosis of syphilis infection in men and nonpregnant women. *Cochrane Database Syst Rev.* (2018) 2018:CD013036. doi: 10.1002/14651858.CD013036

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 da Rocha, dos Santos, Fontes, de Melo, Cunha-Oliveira, Miranda, de Oliveira, Oliveira, Gusmão, Lima, Pinto, Barros and Valentim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fpubh.2022.860396

An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction

Junaid Rashid[1], *Saba Batool*[2], *Jungeun Kim*[1]*, *Muhammad Wasif Nisar*[2], *Amir Hussain*[3], *Sapna Juneja*[4] *and Riti Kushwaha*[5]

[1] Department of Computer Science and Engineering, Kongju National University, Cheonan, South Korea, [2] Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan, [3] Data Science and Cyber Analytics Research Group, Edinburgh Napier University, Edinburgh, United Kingdom, [4] Department of Computer Science, KIET Group of Institutions, Ghaziabad, India, [5] Department of Computer Science, Bennett University, Greater Noida, India

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Praveen Kumar,
VIT University, India
Ramani Kannan,
University of Technology Petronas, Malaysia
Ashwani Kumar Vashishtha,
Vidya College of Engineering, India
Ibrahim El-Bayoumy,
Tanta University, Egypt

***Correspondence:**
Jungeun Kim
jekim@kongju.ac.kr

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 22 January 2022
Accepted: 22 February 2022
Published: 31 March 2022

Citation:
Rashid J, Batool S, Kim J, Wasif Nisar M, Hussain A, Juneja S and Kushwaha R (2022) An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction. Front. Public Health 10:860396. doi: 10.3389/fpubh.2022.860396

Chronic diseases are increasing in prevalence and mortality worldwide. Early diagnosis has therefore become an important research area to enhance patient survival rates. Several research studies have reported classification approaches for specific disease prediction. In this paper, we propose a novel augmented artificial intelligence approach using an artificial neural network (ANN) with particle swarm optimization (PSO) to predict five prevalent chronic diseases including breast cancer, diabetes, heart attack, hepatitis, and kidney disease. Seven classification algorithms are compared to evaluate the proposed model's prediction performance. The ANN prediction model constructed with a PSO based feature extraction approach outperforms other state-of-the-art classification approaches when evaluated with accuracy. Our proposed approach gave the highest accuracy of 99.67%, with the PSO. However, the classification model's performance is found to depend on the attributes of data used for classification. Our results are compared with various chronic disease datasets and shown to outperform other benchmark approaches. In addition, our optimized ANN processing is shown to require less time compared to random forest (RF), deep learning and support vector machine (SVM) based methods. Our study could play a role for early diagnosis of chronic diseases in hospitals, including through development of online diagnosis systems.

Keywords: medical diagnosis, feature selection, chronic diseases, artificial neural network (ANN), prediction

INTRODUCTION

Hospitals and online medical systems are generating a large amount of data that is beneficial for researchers to apply Artificial Intelligence (AI) techniques in building advanced models. Disease diagnosis, disease stage prediction, medical wearable, hospital stay, and death prediction are the main research areas in medical computing. Early disease detection and risk identification can motivate people to change their lousy eating, lifestyle, and exercise habits. Early chronic disease risk identification is beneficial for patients to start treatment early (1). Artificial intelligence in the diagnosis of medical diseases is a trend research area. Artificial intelligence techniques and

algorithms help physicians in the decision-making process. A health specialist typically uses medical lab tests and experiments to decide whether a person is suffering from a specific disease or not. It is now easier to acquire information from previously available data using artificial intelligence techniques. In the present electronic era, massive amounts of data are generated from different health devices, namely sensors, clinical databases, social networks, and wearables (2). Except for laboratory tests and experimental results, age, gender, drug addiction, body mass index (BMI), blood pressure (BP), hypertension, lifestyle, diet, and exercise habits of patients are factors for the most common predictions of chronic diseases, including diabetes, heart attack, hepatitis, and kidney diseases (3). Artificial intelligence techniques are used to predict diseases based on available patient data. Medical diagnosis requires physicians and medical laboratories for testing, while artificial intelligence-based predictive systems are used for the early prediction of diseases.

Artificial intelligence is inspired by biological processes, especially input processing approaches used by the human brain. Artificial intelligence can learn, process inputs, identify patterns, predict, and make decisions (4). Artificial intelligence techniques are mainly divided into supervised and unsupervised learning. Supervised learning needs training on datasets available with required outputs. This training is used to build a model that is applied for future predictions (5). In unsupervised learning, similar data points are combined into groups called clusters. Each cluster consists of data points with similar characteristics (6). Supervised learning is further divided into two categories: Classification and Regression. Classification is used to predict the final class labels with positive and negative results. Classification algorithms are used to build a predictive model based on historical data and indicate classes of unlabeled data.

According to the World Health Organization (WHO) analysis, 41 million people die from chronic diseases, which is 71% of total deaths each year. Chronic diseases include heart attacks, cancer, diabetes, kidney disease, liver disease, etc. Diabetes killed 1.6 million people in 2016 (7). This rapid growth of diabetes patients requires an early detection system to treat their disease early. Diabetes mellitus occurs with an increase in blood glucose levels. There are mainly two types of diabetes: Type I diabetes occurs when the pancreas stops producing the required insulin. Type II diabetes occurs when a patient's body is not responding (8). Diabetes is a lifelong disease that affects different body organs, including the kidneys, heart, and eyes. Heart diseases are considered as another main reason for death in middle and old age. According to cardiovascular statistics provided by the World Health Organization (WHO), more than 17.9 million deaths occur due to heart stroke (9). Physicians refer to medical tests for heart disease diagnosis like electrocardiogram (ETC), echocardiogram (ECG), computerized tomography (CT scan), etc. In another way, early heart attack symptoms are used to predict disease so that the clinical burden can be reduced and physicians or patients themselves can diagnose their heart disease. Family history, blood pressure level, cholesterol, stress level, eating habits, age, weight, smoking, and alcohol consumption are the main factors that are examined to predict heart disease before occurrence.

Conventional methods of kidney disease diagnosis involve ultrasound, urine or blood tests, etc., which are followed by kidney dialysis, transplant, or simple medication. Almost one out of three diabetes patients suffer from kidney failure (10). Kidney diseases are predicted early by using automatic prediction systems. Early detection may reduce the mortality rate and cost associated with lab experiments. Kidney diseases are related to bad diet habits and unhealthy living; these attributes are easily used to build a predictive model. Hepatitis is an inflammation in the liver due to viral infections. It is another rapidly spreading chronic disease with more than 350 million patients (11). The reason behind death is hepatitis shows symptoms like viral infections, namely, fever, fatigue, and liver infections, etc.

Artificial intelligence makes it possible to diagnose Hepatitis C early without clinical expense (12). Breast cancer is overgrowing in European and Asian countries. It has become the most frequent type of cancer, causing the most incredible death rate among women that is 15% (13). Breast cancer diagnosis involves two steps: detection and screening. It is crucial to predict breast cancer early to provide necessary medical services and care. Many researchers contribute to the detection and prevention of breast cancer at an early stage. Breast cancer became the second leading cancer that causes the most deaths among women. It is necessary to propose a predictive analysis model for cancer diagnosis in its early stages. Early diagnosis of breast cancer will lead to early treatment and may reduce the risks of death. Web-based systems can also reduce the cost of lab experiments; therefore, they are beneficial to the patient and medical staff as well. Improvements in medical database systems help to generate a large amount of data that is unable to handle by a human. Artificial intelligence is used to extract patterns in huge datasets and to find the labels of data for patient disease diagnosis (14).

However, health-related data are more sensitive; therefore, they must be handled carefully to generate accurate results. Real-World data has many issues, including noisy data, class imbalance, missing values, and irrelevant entries (15). Different data pre-processing approaches have been applied in previous research to increase the performance of medical diagnosis systems (16). Data pre-processing plays a vital role in extracting valuable features in data. Data pre-processing consists of extraction, cleaning, and transformation steps (14). Chronic diseases are affected by hypertension, gender, obesity, age factors, smoking, unhealthy diet, and lifestyle; therefore, it is difficult to identify the disease from all functional aspects.

Furthermore, different hospitals or medical diagnostic laboratories have the attributes of other patients with the same strategies. A preferable method to handle other characteristics in the dataset is feature selection. Feature selection approaches make it possible to find weights of different factors so that only the most influencing factors are used for detection. Feature selection is used to select the most significant features and to eliminate irrelevant features from the dataset. Feature reduction techniques reduce processing time and increase classification performance that is affected by useless features in the data (17).

Some datasets are available online in different data repositories for research purposes. To find and evaluate the symptoms of other diseases, the researchers applied

TABLE 1 | Summary of the literature review.

Year	Problem statement	Diseases	Proposed solution	Reference
2021	Disease prediction with the features using machine learning	Diabetes and breast cancer	SVM	(77)
2020	Different feature selection approaches are compared to evaluate their recall, precision, and F1 measure performance.	Diabetes, Kidney and Heart Attack	Adaptive probabilistic divergence is used to select most useful features.	(2)
2020	To predict the presence of three chronic diseases.	Diabetes, Heart attack, and cancer	Incremental Feature Selection Approach with Convolutional Neural Network (CNN)	(18)
2019	To explore most important features for different chronic diseases.	Heart, Hepatitis, Diabetes, Cancer	Including information gain, gain ratio, and correlation-based approaches.	(19)
2020	Enhance the accuracy of a prediction model	Chronic Diseases	Stacked Ensemble approach	(20)
2020	To enhance the classification results using the clustering method	Diabetes, Cancer & Kidney diseases	Rough K-means clustering	(21)
2019	Focused on feature filtering techniques to predict cancer in an early stage.	Cancer	Decision Tree, Naive Bayes, k-Nearest Neighbors, and Support Vector Machine	(23)
2020	Extract the most influencing features	Cancer	Pearson correlation with ANN	(25)
2019	Enhance classification performance using significant attributes	Diabetes	A hybrid of AdaBoost, Bagging, and K-NN	(53)
2019	To predict diabetes using demographics and hypertension data	Diabetes	convolution neural network (CNN)	(54)
2020	To highlight features most common in heart and diabetes patients.	Diabetes and Heart Disease	Supervised learning classification and regression	(8)
2019	Heart disease prediction based on patients' data.	Heart Attack	Random Forest	(55)
2019	Feature reduction in heart patients' data.	Heart Attack	Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO)	(56)
2020	To explore the most significant features	Hepatitis	Principal Component Analysis (PCA)	(66)
2019	Highlight hepatitis C risk predicting factors	Hepatitis	Random Forest	(68)
2020	Feature transformation is used to increase the accuracy of the hepatitis prediction model	Hepatitis	Classification	(11)
2019	Irrelevant significant features reduction for kidney disease prediction model	Kidney Failure	Ant Colony Optimization	(17)
2019	to predict chronic kidney disease based on demographic data.	Kidney Diseases	Random Forest	(73)
2020	Ensemble techniques for kidney disease prediction	Kidney Diseases	Decision Tree, k-NN and Naive Bayes	(74)
2019	A prediction model for patients' survival after kidney transplant	Kidney Diseases	Information Gain and Naive Bayes	(75)

various algorithms and built different effective prediction models. These models are applicable for the selected dataset to diagnose any specific disease early. In comparison, many research studies have proposed methods to predict various diseases, e.g., Diabetes, Cancer, Parkinson's, Heart Attack, Depression, Hypertension, etc. However, the medical diagnosis systems focus on a specific disease with limited attributes. Previous studies have shown that traditional methods for predicting chronic diseases are inadequate due to very limited feature selection techniques. Therefore, significant features identification and machine learning techniques are important to improve the performance of chronic diseases. The main contributions of this research paper are as follows.

- The significant features are chosen to eliminate redundant and inconsistent data through particle swarm optimization, which improves the efficiency of prediction model.
- An artificial neural network predicts five chronic diseases, diabetes, breast cancer, hepatitis, heart, and kidney.

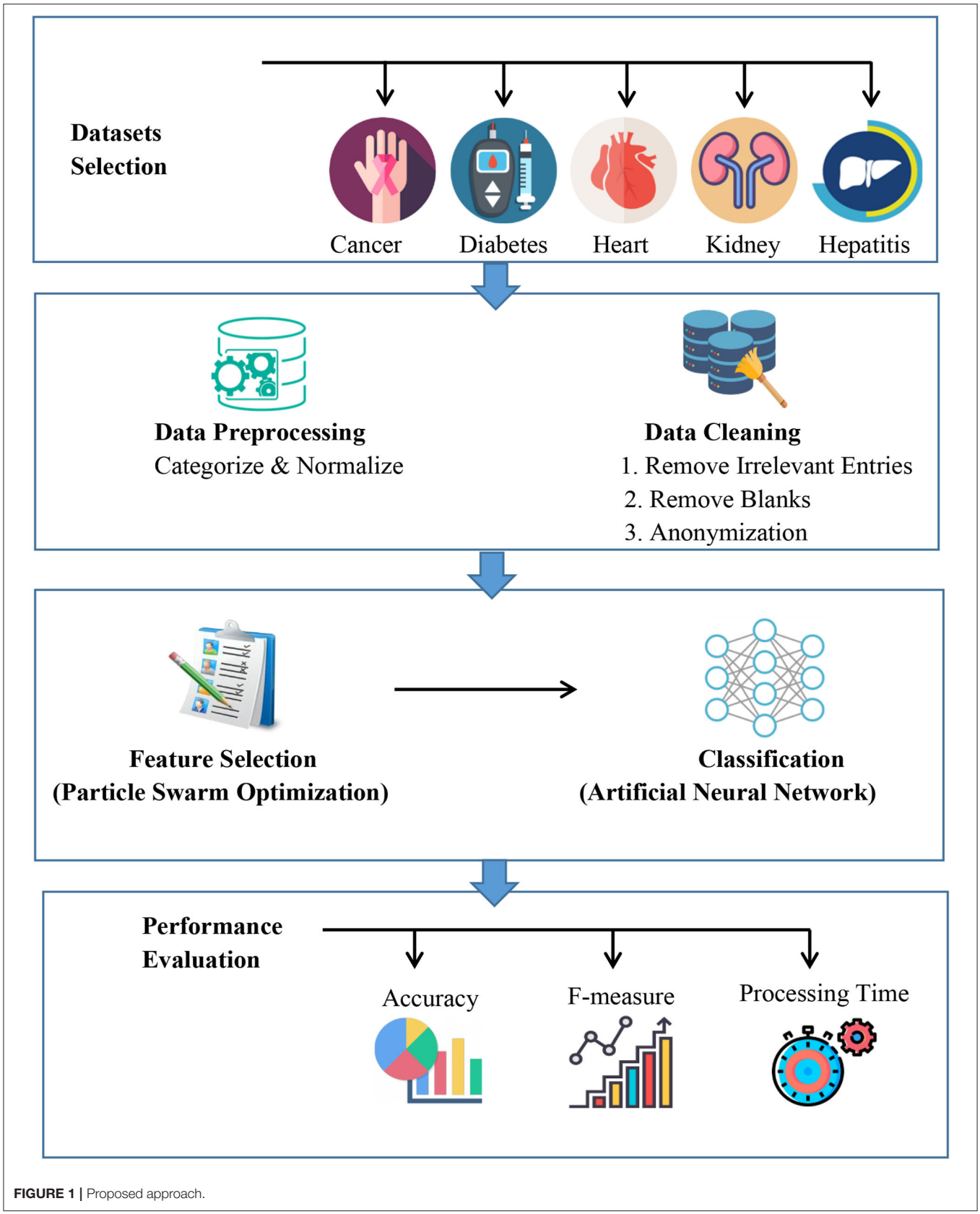


FIGURE 1 | Proposed approach.

TABLE 2 | Datasets used for classification.

Diseases	Dataset name	No. of attributes	Total entries	No. of positive results	No. of negative results	References
Cancer	Breast Cancer Wisconsin (Diagnostic) Data Set (Dataset 1)	31	569	212	357	(78)
Cancer	Breast Cancer Wisconsin (Dataset 2)	10	699	240	459	(79)
Diabetes	Pima Indians Diabetes Database (Dataset 1)	9	768	268	500	(80)
Diabetes	Diabetes Classification (Dataset 2)	15	390	60	330	(81)
Heart Attack	Heart Disease UCI (Dataset 1)	14	303	138	165	(82)
Heart Attack	Heart Disease Prediction (Dataset 2)	13	270	150	120	(83)
Hepatitis	Hepatitis (Dataset 1)	19	142	80	62	(84)
Hepatitis	Indian Liver Patient Records (Dataset 2)	11	584	168	416	(85)
Kidney	Kidney Disease Dataset (Dataset 1)	26	189	74	115	(86)

- The k-fold cross-validation is used to train and test the proposed approach. Classification results are compared to a decision tree, random forest, deep learning, KNN, Naive Bayes, SVM, and logistic regression on chronic diseases datasets.
- The experimental results show that the proposed approach achieved a 99.76% accuracy and predicts chronic diseases more effectively and efficiently than other state-of-the-art models.

The rest of the study is arranged as follows: Section Literature Review represents a comprehensive review of studies in the previous literature. Section Materials & Methods discusses the materials and methods. Section Results and Discussions presents experimental results and comparative analysis. Finally, Section Conclusion concludes, and section Future Work and Limitations provides future directions and limitations for the research.

LITERATURE REVIEW

The literature review provides a clear view of previous approaches and potential areas. The latest literature reviews different disease detection methods using artificial intelligence approaches with or without other feature selection methods.

A research study proposed a machine learning-based predictive model to diagnose three primary chronic diseases, including Diabetes, Kidney, and heart diseases. The feature selection technique based on adaptive probabilistic divergence is used to select the most valuable features. The study concluded that the proposed approach gave the highest accuracy by optimizing the most significant features for disease detection (2). A feature selection-based machine learning algorithm is proposed to predict three chronic diseases, namely, diabetes, heart attack, and cancer. The incremental feature selection approach with Convolutional Neural Network (CNN) is applied to anticipate disease presence. The proposed method showed 93% classification accuracy in less computation time (18). Another study is conducted to explore the essential features of often chronic diseases. Feature selection approaches, including Information Gain, Gain Ratio, and correlation-based approaches, are applied. Several subsets of top-ranked features are then used in building the Random Forest prediction model. It showed that exploring the most significant features is significant for the medical diagnosis process (19). In another research study for chronic disease prediction (20), the Stacked Generalization approach is used to enhance the performance of classification algorithms. Five classification algorithms are compared: Decision Tree (DT), k-NN, SVM, Logistic Regression (LR), and Naive Bayes, to outperform five chronic disease prediction models. It is found that the Stacked Ensemble approach enhances the model performance and achieves the highest accuracy of 90%. Diabetes, Breast Cancer & Kidney diseases are predicting in (21), using classification algorithms. Rough K-means clustering is used to explore and eliminate ambiguous objects in datasets. It is observed that classification model applied with extracted features gave improved results compared to conventional models.

Early breast cancer prediction is an emerging research area in artificial intelligence. Many studies are conducted for breast cancer prediction at an early stage using online techniques and prediction models. The fuzzy rule-based classification is used for classification purposes, while fuzzy temporal rules are used to determine highly contributing factors for online breast cancer prediction. The results show that feature selection and fuzzy rule-based classification improve the accuracy of classifiers (22). Another study focused on feature filtering techniques to predict breast cancer early. Frequent item-set mining is used to select the essential features in patients' datasets. The decision tree, Naive Bayes (NB), k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM) are compared, and it is found that SVM outperforms other models (23). A research paper focused on reducing erroneous prediction results that is false positive and false negative. Information gain with the Genetic Algorithm approach is utilized to rank highly significant features in the dataset. Classification of positive and negative results is done by SVM. This approach applied to

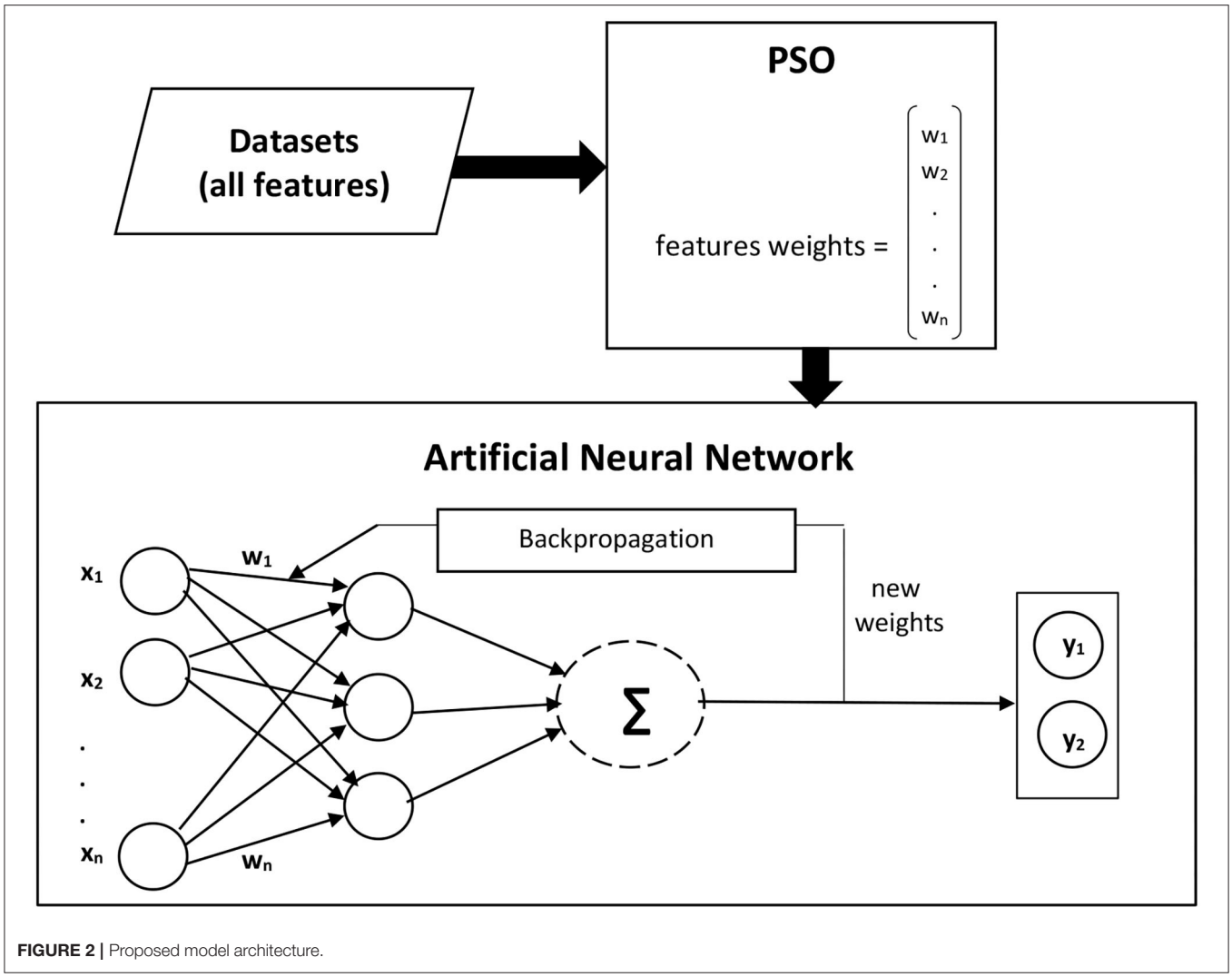


FIGURE 2 | Proposed model architecture.

two independent datasets increased the prediction accuracy and reduced prediction cost (24). In another research study (25), five classification models, namely SVM, K-NN, ANN, random forest, and logistic regression, are applied to predict early breast cancer. In this study, the Pearson correlation is applied to extract the most influencing features, and after feature selection, the ANN model obtained the highest accuracy of 98%. Two breast cancer datasets are used in (26), to differentiate cancer patients from healthy people. A genetic algorithm is used to select the most significant features in datasets.

Three classification algorithms, multilayer perceptron (MLP), probabilistic neural network (PNN) and radial based function (RBF) are used for the classification of breast cancer (27). MLP requires more processing time for the training model and assigning weights to its neurons than RBF and PNN. The highest accuracies of 97%, 98%, and 100% are achieved with MLP, RBF, and PNN. Another research study (28) applies SVM with multiple kernels to predict breast cancer. An energy-based shape histogram is used to extract the most significant features. The proposed model achieved the highest accuracy of 99%.

Diabetes prediction at an early stage can lead to saving patients' life. A research paper compares classification methods like Artificial Neural Network (ANN) and Random Forest with clustering technique k-means clustering. Principal Component Analysis (PCA) is used to select significant components. The study proved that PCA increases the accuracy of the prediction of diabetes, while body mass index (BMI) and glucose level strongly correlate with diabetes (28). Prolonged diabetes can cause vision loss, which is called diabetes retinopathy. A research study is conducted to predict the side effects of diabetes on eyesight. K-NN, Decision Tree, Multilayer Perceptron, and SVM are used to predict feature selection. The number of accurate predictions is compared before and after feature selection, demonstrating that feature selection approaches increased accuracy and sensitivity (29). Some techniques are used for diabetic retinopathy using PCA (30) and deep neural network (31). In (32), BMI is predicted from the images using the computer vison and

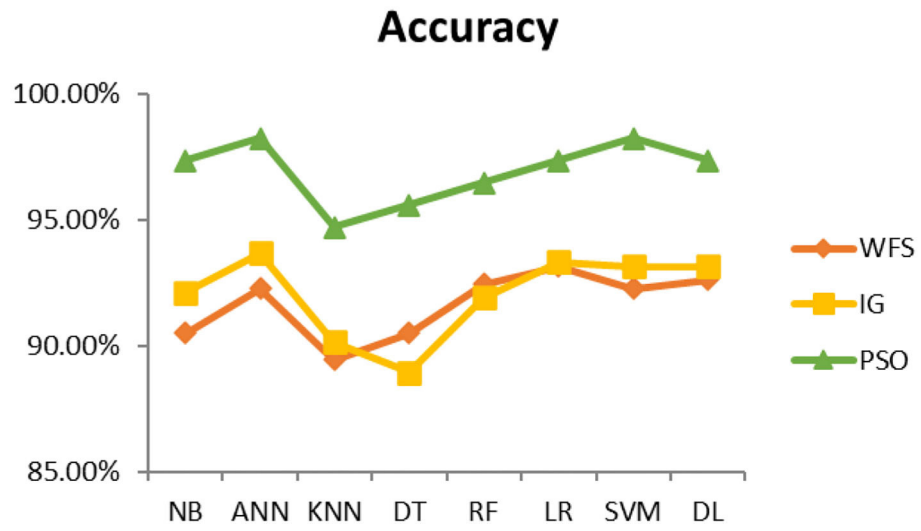


FIGURE 3 | Cancer diagnosis accuracy (dataset 1).

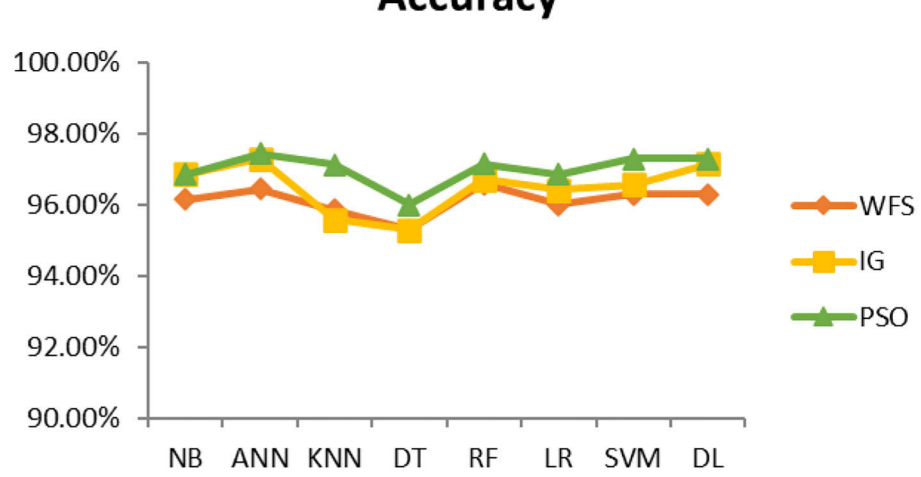


FIGURE 4 | Cancer diagnosis accuracy (dataset 2).

machine learning technique. Federated learning has a wide application in big data (33). The probabilistic procedure with sensors is used for censorious procedures (34). Some patients have diabetes with hypertension; therefore, a research study is conducted to predict diabetes type II and hypertension in both individuals. Synthetic minority oversampling is used to solve data distribution problems and the ensemble method is used to predict hypertension and type II diabetes. This study showed that pre-processing of data prior to model building enhances prediction accuracy (35). Machine learning methods are also used for the detection of some other medical diseases detection (36–38), text mining (39–42) and network security (43–47). Another analysis is conducted to predict the risks associated with diabetes. Logistic regression, Decision Tree, ID3, C4.5, k-NN, and Naive Bayes are used for classification. Irrelevant features are detected and reduced using PCA and PSO algorithms. A comparison of both feature selection techniques is performed in terms of improved accuracy and processing time. It shows that feature reduction is a powerful technique to enhance predictive model accuracy (48).

Iris-based and physiological features are used to predict diabetes type II, using decision trees and SVM. Some methods used the fuzzy for ergonomics-related disorders (49), Convolutional Neural Networks for Syndrome (50), object detection (51) and gender classification (52). Three ensemble techniques, AdaBoost, Bagging, and K-NN, are also used with Principal Component Analysis (PCA) to enhance classification performance. The highest accuracy of 95.81% is obtained

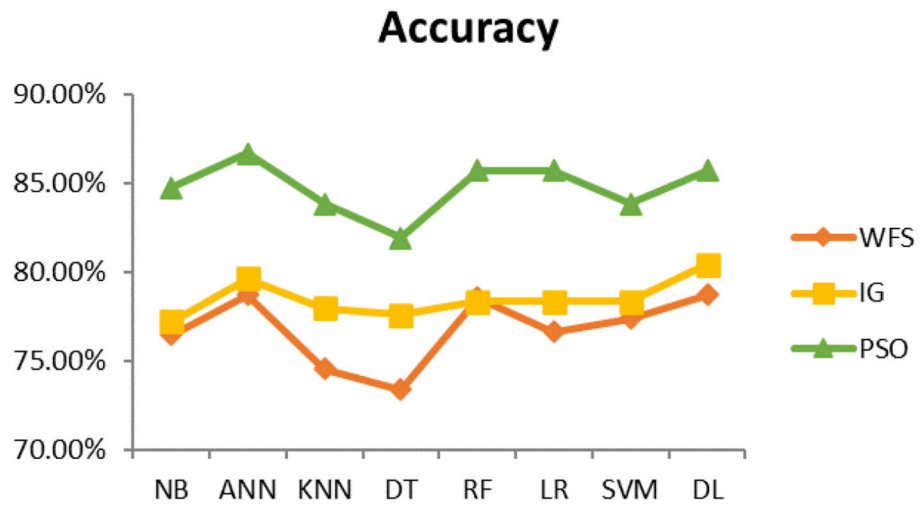


FIGURE 5 | Diabetes diagnosis accuracy (dataset 1).

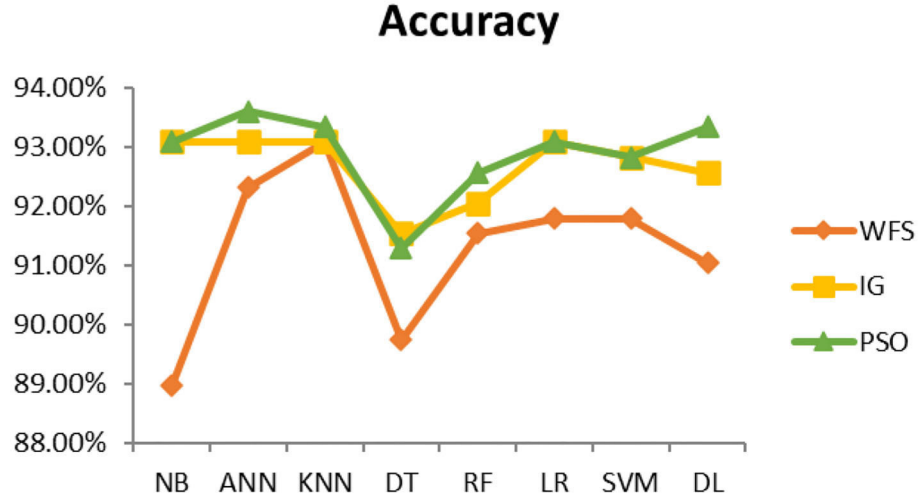


FIGURE 6 | Diabetes diagnosis accuracy (dataset 2).

with the most significant features (53). A convolution neural network (CNN) is a feed-forward and feature extractor Neural Network. In a study (54), early diagnosis of Diabetes Type II is performed by using CNN. A dataset of steel mill workers, including demographics, physical activities, and hypertension features, is used to implement CNN. The highest accuracy of 94.5% is obtained to diagnose diabetes patients from all workers' datasets. Diabetes patients are usually affected by some other chronic diseases. A research study (8) highlighted the features most common in heart and diabetes patients. Heart disease is found to be predicted early in diabetes patients by applying classification and regression techniques.

In recent decades, heart attacks have been one of the leading causes of death. Random forest is used in conjunction with a linear model to predict heart disease based on patient data in a study. The most significant features are classified using an a priori method, which showed an increase in accuracy. When compared to existing classification techniques, the suggested ensemble model produces the best results (55). Optimization techniques are used to deal with complex data by reducing irrelevant and functional data attributes. In a study, the Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) algorithms are combined for redundant feature reduction. This approach is applied with different classification algorithms and is evaluated using F-measure, accuracy, precision, and recall. Feature optimization gave the highest accuracy of 99.65%, which is a remarkable achievement (56). Software quality (57–59) and semantic methods (60, 61) are also used in algorithm evaluations. A hybrid approach combines different AI techniques and evaluates them as a single algorithm. A research study

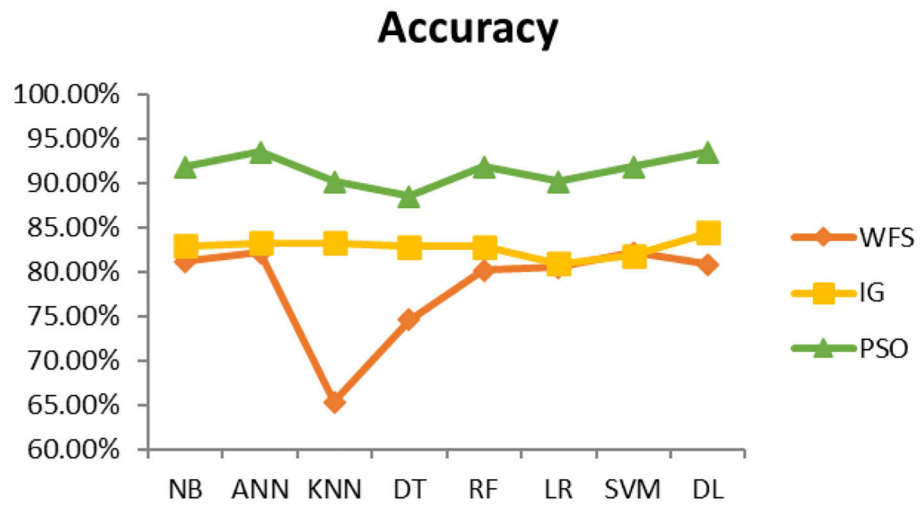


FIGURE 7 | Heart attack diagnosis accuracy (dataset 1).

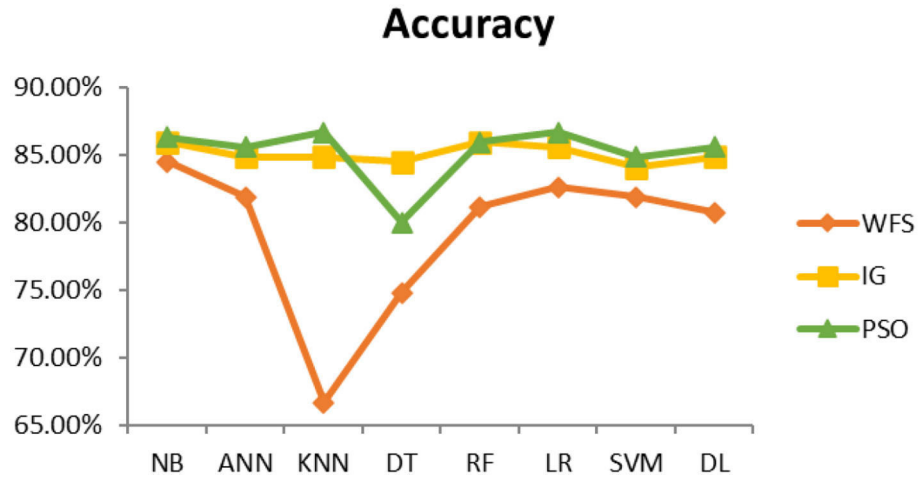


FIGURE 8 | Heart attack diagnosis accuracy (dataset 2).

applied various AI algorithms for heart disease prediction to find the best classifier. The best performing algorithms are then combined, and the results of the hybrid approach are compared with the results of individual techniques. The proposed hybrid algorithm gave more accurate results as all algorithms functions are combined to calculate a majority vote for a positive class. A drawback of using a hybrid approach is that more processing time is required (62). Another study applied a Fuzzy Rule-based classification model for heart attack risk prediction. Rough set theory identifies the most influencing features, and features that do not affect prediction results are eliminated. The Genetic Algorithm is used to reduce time complexity and optimize prediction results. Research proved that the proposed model is efficient with a large number of features (63, 64). A comparative analysis is presented in (65) for the diagnosis of cardiovascular disease. The experimental setup comprises classification algorithms, namely Logistic Regression, Decision Tree, and SVM. Different subsets of features are used to evaluate classification results. The highest accuracy of 82.97% is achieved by using a decision tree with forward and backward selection approaches.

A study is proposed to predict whether a patient would survive or die because of hepatitis. This study implemented SVM with multilayer perceptron for the binary classification of patients. Principal component analysis (PCA), a feature selection approach, is applied before prediction and three factors of zero correlation with the predictive class are eliminated. The proposed hybrid algorithm gave a higher accuracy than state-of-the-art classifiers (66). A combination of several prediction techniques results better than an individual model. Another study suggested a model for hepatitis diagnosis using ensemble learning techniques. A Neuro-fuzzy system

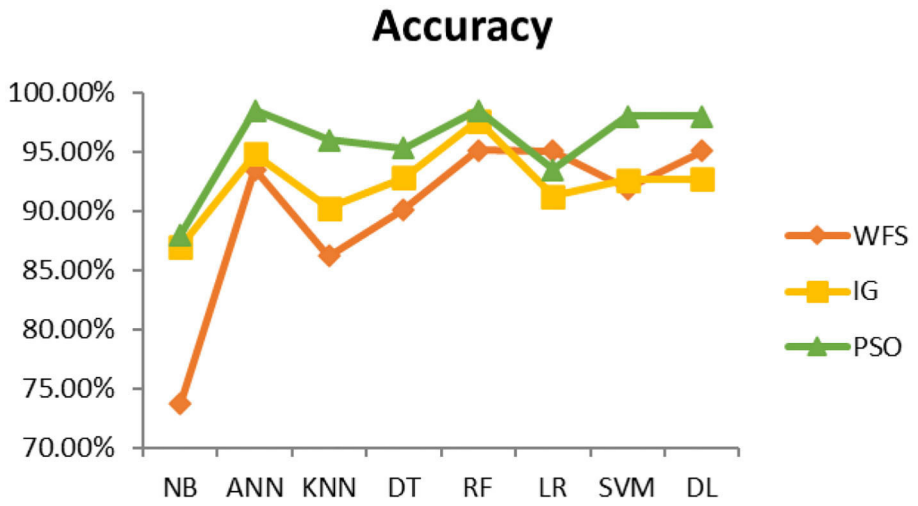


FIGURE 9 | Hepatitis diagnosis accuracy (dataset 1).

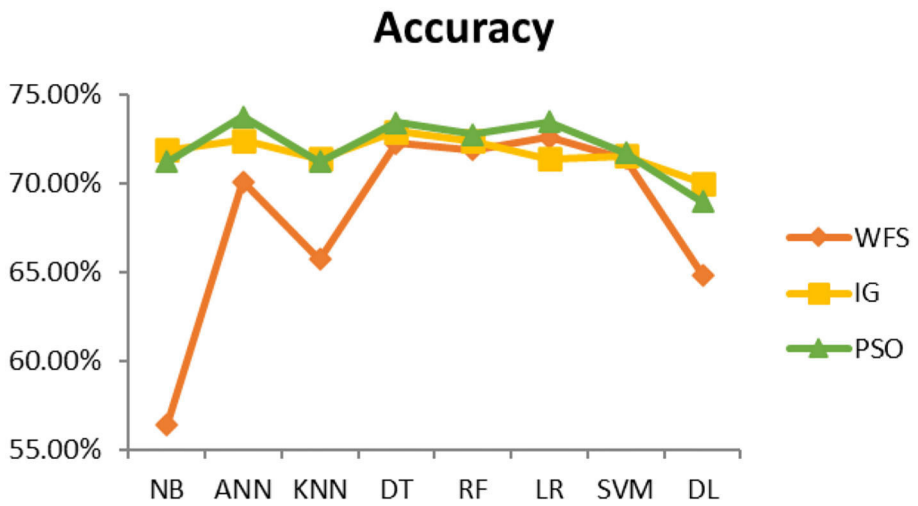


FIGURE 10 | Hepatitis diagnosis accuracy (dataset 2).

is used for patients' diagnosis with dimensionality reduction algorithms. The results demonstrate that the ensemble neuro-fuzzy model performed more accurately than ANN (67). A study is conducted to highlight risk prediction factors for hepatitis C using a random forest classifier. Chi-square and *P*-value statistical functions are used to estimate the importance of various risk determining factors. The importance of the factor is calculated several times and each important timeless factor is eliminated until the required predictive model has 98.3% accuracy. The research found that gender, drug addiction, frequent hospitalization, and migration from the area with a high HIV occurrence rate are the essential factors for hepatitis diagnosis (68). A two-step feature transformation method is proposed to increase the accuracy of the hepatitis prediction model. At first, the most valuable features are ranked based on their significance, and after feature reduction, the attribute space is expanded. This hybrid method manifested improved prediction results achieved in less computation time (11). Esophageal variants occur most commonly due to the side effects of hepatitis. The esophageal prediction model is proposed in a research study, with six feature selection approaches. Information Gain, PCA, and correlation are used to select the most significant features. The Greedy Approach, PSO, and Genetic Algorithm are applied to create subsets of the dataset by exploring the correlation between different features. This study concluded that Naive Bayes outperformed compared to ANN, Decision Tree, Random Forest, and SVM (12). Some other machine learning and soft computing methods are mainly used for the chronic disease detection (21, 69–71).

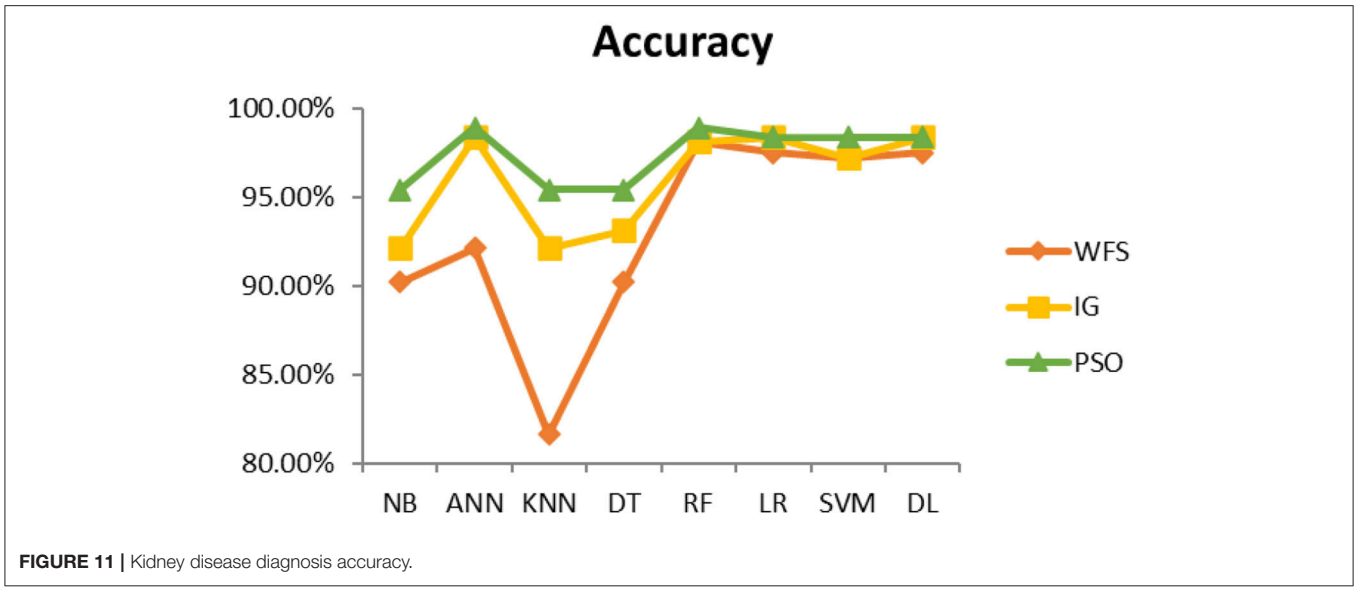


FIGURE 11 | Kidney disease diagnosis accuracy.

TABLE 3 | Highest accuracy achieved by proposed model.

Diseases	Highest results achieved by proposed model
Cancer	98.23%
Diabetes	93.59%
Heart	93.44%
Hepatitis	98.46%
Kidney	98.90%

Ant Colony Optimization is used for reducing irrelevant or least significant features. The study presented that ACO feature selection enhances the performance of supervised learning classifiers (17). Kidney disease stage identification is performed by applying artificial intelligence techniques to various patient attributes. The most influencing features in the prediction of kidney disease stage are serum creatinine, urine, and albumin. Probabilistic Neural Network beat other prediction models with 96.7% accuracy (72). Random Forest is applied to predict chronic kidney disease in patient data. Dataset is divided into demographics and clinical data, including patients' age, BMI, gender, diabetes history, hypertension, etc. Prediction results are estimated by R2 (73). Data are preprocessed by outlier detection, normalization, and irrelevant data features are reduced. The random subspace outperformed with 100% accurate results and proved that ensemble approaches are better to use than individual classifiers (74). Besides kidney disease presence prediction, kidney transplantation survival is another critical aspect of prediction. The study proposed a prediction model for patients' survival after kidney transplants. Information Gain and Naive Bayes are used for the most significant feature selection. When evaluated with Accuracy and F-measure, K-NN gave higher accuracy than other classifiers (75). Another study classifies kidney patient and healthy people data by applying correlation-based feature selection approaches. Three feature ranking approaches, namely ReliefF, information gain, and gain ratio, explore the most significant attributes (76). **Table 1** shows the summary of previous literature that presents their problem statements, diseases to be predicted, and proposed solutions.

MATERIALS AND METHODS

In this section dataset selection, data pre-processing, feature selection, proposed model architecture, classification methods, performance evaluation, experimental tools and setup and validation method are discussed. **Figure 1** describes the proposed automated artificial intelligence approach. The datasets are selected, and pre-processing is applied. The features are extracted with particle swarm optimization and then an artificial neural network is utilized.

Dataset Selection

Datasets for five chronic diseases, including Breast Cancer, Diabetes, Heart, Hepatitis, and kidney disease, are collected from multiple online sources e.g., Kaggle, Dataworld, Github & UCI machine learning repository. Four datasets are previously used in a research study (19), to predict different diseases, while other datasets are used in multiple research studies. Each dataset has information about numerous patients with various diseases; therefore, dataset features and instances vary. All datasets used in this research study have numeric and categorical data features based on patients' demographics and medical test results. Breast Cancer data attributes include the characteristics extracted from digital images of fine-needle aspirates of different breast tissues. Diabetes datasets are extracted from patients' demographic profiles and medical test results, including gender, age, glucose level, blood pressure, and body mass index, etc. Two datasets containing heart disease symptoms and patients' data are used to detect a person's heart disease or not. Heart disease dataset attributes used for detection include demographics, chest

pain, cholesterol level, glucose level, ECG results, maximum heart rate, depression, and exercise-induced angina. For the early prediction of hepatitis, demographics and medical lab test attributes are used. Dataset features used for diagnosis are categorical, including age, gender, antivirals, liver size, fatigue, discomfort, bilirubin, albumin, etc.

The kidney disease dataset is used to diagnose kidney disease presence, using features like sugar level, red and white blood cell count, glucose level, blood urea, diabetes presence, and hunger. All above datasets consist of multiple features; however, not all features are equally significant for disease diagnosis. This research study finds that feature selection approaches help to optimize classification results by considering useful features only. Therefore, for each disease, two publicly available datasets are used to verify the classification results. However, for kidney disease classification, only one dataset is found on multiple online platforms. **Table 2** describes the dataset attributes and their statistics.

Data Pre-processing

In predictive analytics, data quality is a significant factor, low quality of data leads to incorrect results. Due to their varied uses and sources, real-world medical data sets have many outliers, missing values, and irrelevant features. To make the artificial intelligence algorithm more productive, pre-processing is essential before using data for model training (87). Data pre-processing is a conventional artificial intelligence technique to convert raw data into useful information. Data pre-processing involves various steps needed to obtain relevant and valuable data from the original dataset. Data pre-processing steps involve data extraction, anonymization, integration, cleaning, outlier detection, and duplicate removal (15). Therefore, all the datasets are pre-processed before using data for prediction. In these datasets, it is observed that the patient's glucose level, BMI, skin thickness, insulin, blood pressure, and age have a minimum value of zero, which is impossible. Such outliers are replaced with average values. Some instances have blank entries, which are removed by using the replace missing values operator to make an accurate detection. **Table 2** presents the datasets and their values.

Feature Selection

Many datasets for all diseases are available in the hospital's database systems and online repositories. These datasets have different features for multiple usages; not all available features contribute to specific disease detection. Data pre-processing is an essential step before applying a model for the classification of data. The model trained on a dataset with wrong entries may lead to misleading results. Similarly, not all attributes in a dataset contribute to the detection process equally. Including irrelevant features increases the processing time of the model and it also decreases model performance. If all attributes are used for predictive analysis, classifier performance is affected negatively. For hospitals and medical labs, it is dependent on the physician's experience to diagnose a disease from different symptom input values. However, different feature selection approaches allow examining the contribution of various features on results. The main functionality of feature selection techniques is used to analyse that how each attribute takes part in output prediction. Feature selection is an essential approach to be used before model construction to redue data complexity by eliminating irrelevant and useless features from data. Feature selection approaches reduce the data size so that the model takes less time in the training and testing phases. Feature selection is beneficial as it reduces the data size so that processing time, space, and power consumption are also reduced (15). With a fewer number of features, the classification model becomes more interpretable and yields more effective results. Features with zero or very low contributions are eliminated by using feature selection. There are various feature selection methods. Information Gain (IG) is one of the most effective methods for feature selection. Information Gain is measured by entropy, which is the uncertainty of being selected. A low entropy value represents more chances of being selected as a final class label (16). Information Gain calculates the weight of each attribute to estimate its contribution to the last class selection. The highest weight represents the feature that has the most information about final class selection, therefore considered as the most influencing feature.

We present the significance of selecting the most significant features using Particle Swarm Optimization (PSO). PSO results are compared with Information Gain to evaluate its performance. Particle Swarm Optimization (PSO) is another feature ranking technique. PSO is introduced in 1995, inspired by groups of birds and fishes searching for food in a swarm group. Each species in the swarm searches for food, and the whole group follows that individual nearest to the food to travel in the right direction in minimum time. Similarly, the PSO technique is used to find an optimal solution for a problem consisting of various features (48). PSO is an evolutionary algorithm developed on the social behaviors and movements of species. PSO is applied to optimize classification results by assigning weights to the most significant features based on their previous results (88). Thus, only the significant features are used for the prediction process to decrease processing time and to enhance the efficiency of prediction model. The ANN classification model based on the most significant features obtained by PSO. Using PSO, all attributes are assigned weights according to their contribution to the classification process. In addition, a baseline information gain approach is also used to compare its results with the proposed model.

Proposed Model Architecture

The proposed model architecture is described in this section. Artificial intelligence algorithms are used to construct the chronic disease detection model. Artificial intelligence is an advanced machine learning approach to build detection models that can receive input data, train a model, and predict the output of future data. These model-building techniques are divided into two main categories: supervised learning and unsupervised learning. Supervised learning models are provided with a set of input data with desired output labels to construct a predictive technique for future datasets. Supervised learning algorithms include a set of tree-based, instance-based, probability-based, and ensemble algorithms (89). Classification is a supervised machine learning approach that works under supervised learning to predict the

TABLE 4 | Overall accuracy achieved for all diseases prediction.

Model	Cancer	Cancer	Diabetes	Diabetes	Heart	Heart	Hepatitis	Hepatitis	Kidney	Average
ANN	98.23%	97.43%	86.67%	93.08%	93.44%	85.56%	98.46%	72.73%	98.90%	91.61%
RF	96.46%	97.14%	85.71%	93.59%	91.80%	85.93%	98.46%	72.74%	98.90%	91.19%
DL	97.35%	97.28%	85.71%	93.33%	93.44%	85.56%	98.00%	68.97%	98.38%	90.89%
LR	97.35%	96.86%	85.71%	91.28%	90.16%	86.67%	93.46%	73.44%	98.38%	90.37%
SVM	98.23%	97.28%	83.81%	92.56%	91.80%	84.81%	98.00%	71.70%	98.38%	90.73%
KNN	94.69%	97.13%	83.81%	93.08%	90.16%	86.67%	96.00%	71.19%	95.40%	89.79%
NB	97.35%	96.86%	84.76%	93.33%	91.80%	86.30%	88.00%	71.20%	95.40%	89.44%
DT	95.58%	96.00%	81.90%	92.82%	88.52%	80.00%	95.30%	73.42%	95.40%	88.77%

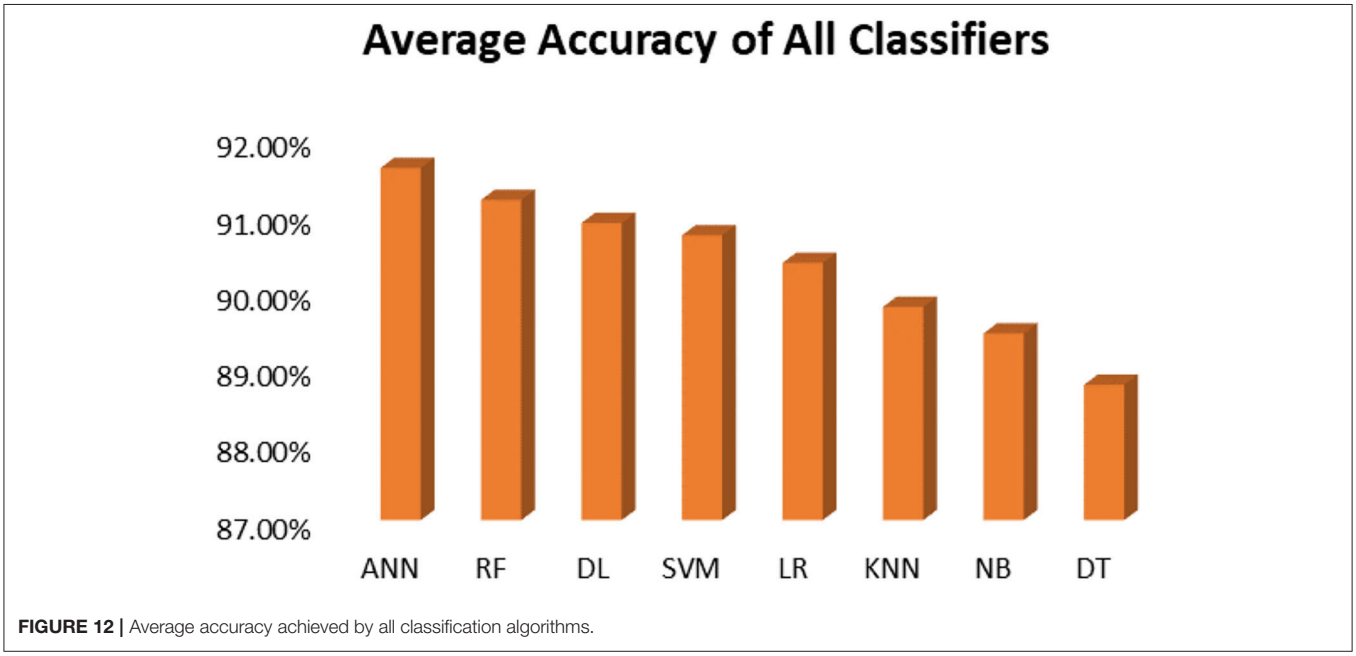


FIGURE 12 | Average accuracy achieved by all classification algorithms.

final class. In classification, the output variable is nominal such as patients' disease diagnosis that is positive or negative (5).

The forward and backpropagation are evaluated during the ANN learning phase, and it is observed that forward propagation gives higher accuracy. Therefore, an Artificial neural network with forwarding selection is used to predict disease presence. Dataset features with and without feature selection approaches are given as an input to the first layer. The processed output of the first layer is fed into the second layer as input until it reaches the output layer. Cross-validation with tens folds is used for the training and testing process. Each dataset is divided into 10 folds, and nine folds are used for training and one-fold for the testing process in such a way that each fold must be used for testing. The different number of training cycles is applied to get the best classification results, and finally, ANN with 200 training cycles is used to finalize the prediction results. ANN learning rate controls how fast Neural Network will learn a prediction problem. It is a configurable hyper-parameter tuned during the ANN training process. A learning rate of 0.01 is used, as a lower learning rate allows Neural Network to learn more optimally. For the breast cancer diagnosis, 31 input features are fed to the proposed model. After classification of the first breast cancer Wisconsin (diagnostic) dataset, the second breast cancer Wisconsin dataset with 10 features is provided as an input to the proposed model, and the results are recorded. Similarly, for the other four diseases detection, datasets with different features, as described in **Table 2**, are fed into the proposed model, and results for each disease diagnosis are recorded. For classification, at first, ANN is trained with all available features to classify patients and healthy people in each dataset. In the second step, features are assigned weights using Information Gain, and classification is made based on the weighted features. In the third step, PSO is applied to each dataset, and its results are compared with classification performance without feature selection and with Information Gain. **Figure 2** presents the architecture of the proposed model, where w stands for the weights assigned to the features, while x and y are used for input and output, respectively.

Classification Algorithms

The Artificial neural network classification algorithm is compared with seven state-of-the-art classification algorithms, namely, Naive Bayes, K-NN, Decision Tree, Random Forest,

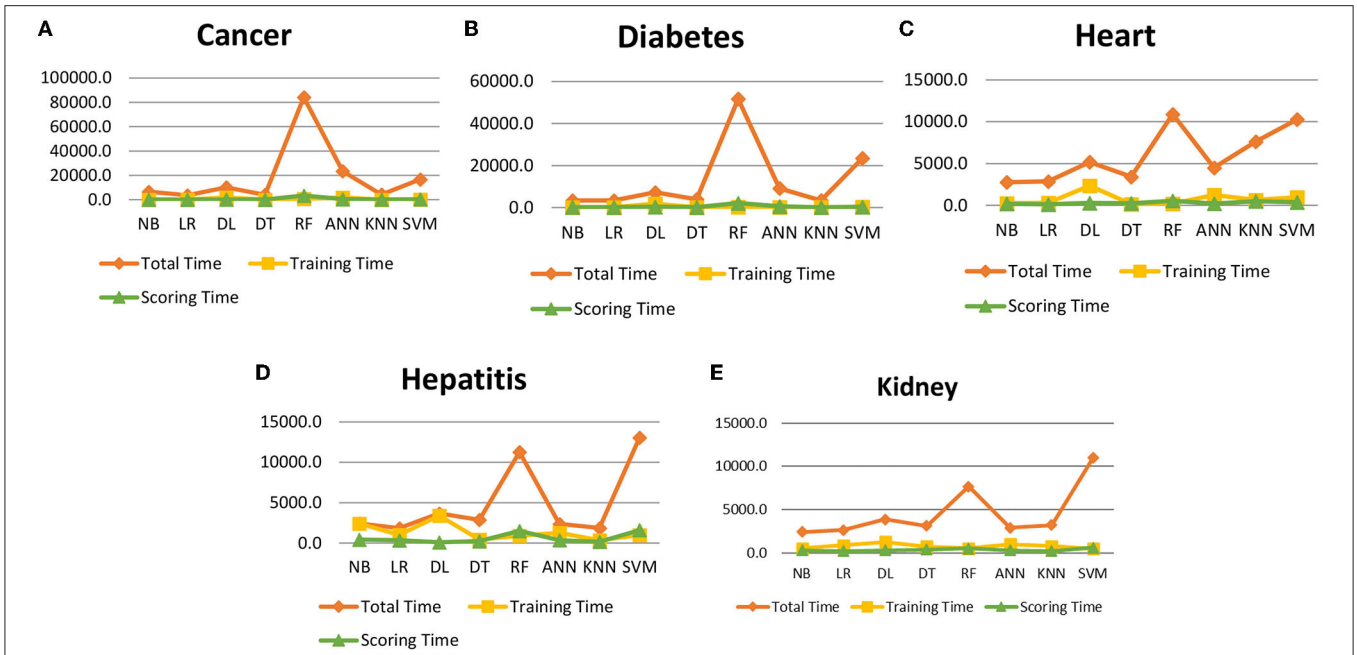


FIGURE 13 | Data processing time for cancer, diabetes, heart, hepatitis and kidney disease.

TABLE 5 | Comparative analysis.

References	Year	Predicted diseases	Classification algorithms	Feature selection	Highest results
Alam et al. (19)	2019	Cancer, diabetes, BP, hepatitis, heart, Parkinson's & carcinoma	Random Forest	Information Gain, Gain Ratio & Relief	99%
Atallah et al. (75)	2019	Kidney transplantation	K-NN	Information Gain & Naive Bayes	81%
Hegde et al. (2)	2020	Kidney, diabetes & hepatitis	Logistic Regression	Probabilistic Feature Selection	91.6%
Sandhiya et al. (18)	2020	Breast cancer, diabetes & heart	CNN	Incremental Feature Selection	93%
Arumugam et al. (77)	2021	Diabetes and breast cancer	Decision tree	Probabilistic Feature Selection	90%
Proposed	2022	Diabetes, breast cancer, hepatitis, kidney & heart diseases	ANN	PSO	99.67%

Logistic Regression, SVM, and Deep Learning. There are various classification algorithms to find a correlation between patients' attributes and disease symptoms, but all algorithms have some limitations in input requirements, model construction function, and required computation time (15). ANN is used for five different chronic disease detection. It is inspired by biological neural networks that pass information from the human brain to other body parts by passing through neurons. ANN is a classification algorithm that learns from training examples and applies the trained model to future datasets to predict their output (16). ANN is a set of connected neurons comprising input, processing, and output layers. Each neuron is connected to its adjacent neurons (90), whereas each connection is assigned some weight. These weights are adjusted in the training phase to weaken or strengthen the relationship based on its contribution toward output prediction. Model training is done through backward or forward propagation (91). Naive Bayes is suitable for both numerical and nominal data with multiple dimensions. Naive Bayes is the Bayesian theorem-based classification algorithm, which works on conditional probability, the class with high probability is considered as the final output. Naive means each feature in the dataset is independent of other features and their presence does not affect others. Naive Bayes can be applied to large datasets as its feature independence makes classification faster than other classification algorithms (16).

KNN is the simplest and easiest algorithm used for classification. K is the nearest neighbor of a data sample whose class is being predicted (91). KNN, also referred to as lazy

learning (15), allocates weights to the contributing neighbors so that the nearest neighbor will have more influence on prediction. Most commonly, Euclidean distance is used to calculate the distance of all neighbors. KNN assumes that similar instances occur close to each other; hence, the nearest neighbors contribute to predicting the final class (75). KNN with a weighted vote is used for classification in this research study, with $k = 10$ which represents 10 nearest neighbors. A Decision Tree is the most straightforward classification algorithm, which follows a tree-like structure starting from the root node and reaches the output node (leaf) by evaluating some conditions. Each node is connected to its next node by a connection (branch); this connection represents a condition, if true, then reaches the next node. Since the decision tree works well with categorical data, it is most suitable for the dataset with fewer features (5). This paper uses decision trees with a maximum depth of 10 nodes for predictive analysis. Numerical data is changed to categorical data by using numerical with the nominal operator. Feature selection techniques are applied to select the most significant nodes so that the classification becomes more effective with a smaller number of features.

Random Forest is an ensemble learning technique used for prediction. Random Forest is a more effective algorithm than a decision tree as it combines a set of multiple decision trees by selecting random data samples and predicts the final class based on the majority vote. More trees mean more robust prediction, but implementing many trees requires more processing time (16). The primary aim of ensemble techniques is to strengthen weak algorithms to build a robust predictive model. In this paper, Random Forest with 100 trees is constructed. Logistic regression is a supervised learning classification technique used to output binary classes. It classifies the output by using probability function, results higher than 0.5 are classified as positive while <0.5 are predicted as negative. It operates numerical input data; therefore, categorical attributes are changed to numerical by using nominal to numerical operators (15). SVM is used to classify linear as well as multidimensional data. In SVM, a hyperplane is used to separate different classes of variables. Both sides of the hyperplane represent two other classes. The Euclidean function is used to calculate the distance between variables so that the hyperplane is drawn in its best place (12). SVM is used for both classification and regression problems. In this paper, SVM with the kernel (dot) is implemented to classify data into binary classes. Before the model, nominal building data is converted to numerical. Deep Learning (DL) is a machine learning approach (92), comprised of a Neural Network with multiple hidden layers. The label with the highest value is finalized as output class (93). Since more training gives more accurate results, but it also increases the processing time and makes the model more complex. Deep learning with epoch 10.0 is applied in this research study.

Performance Evaluation

After training and constructing a prediction model, its performance is tested on similar features with an unseen output class to evaluate how accurately the model performs. A confusion matrix is derived to check the similarity between the actual output and predictions of the designed model. In the confusion matrix, true positive (TP) indicates that positive output is predicted as positive, false positive (FP) indicates that negative class is predicted as a positive class, true negative (TN) represents the true prediction of negative class, and false negative (FN) shows that a negative class is falsely predicted as the positive class. These are the primary measures for any classification model. One of the most used performance measures is accuracy. It is the ratio of the correct number of predictions to the total number of instances. Equations 1, describe the accuracy formula.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{All}} \tag{1}$$

Experimental Tool and Setup

In this research study, RapidMiner is used to implement classification algorithms. RapidMiner is a user-friendly platform for the implementation of artificial intelligence algorithms to solve data analysis and data mining problems. It is open-source software with multiple license types for user convenience. However, it is freely available for academic use. RapidMiner contains hundreds of classification, regression, clustering, and feature ranking operators that are used without coding skills (94). RapidMiner is designed for data mining, data pre-processing and visualization, statistical analysis, feature extraction, prediction, and classification. RapidMiner is easy to use as it has input, output, and processing operators that need drag and drop functions and connection building simply. RapidMiner is widely used for machine learning, text mining, sentiment analysis, future predictions, and statistical data analysis. Its Java-based graphical user interface allows business persons and researchers to use it for data preparation, processing, and business future predictions (95).

K-Fold Cross-Validation

K-fold cross-validation is a technique for dividing data into subsets for training and testing of the model. This technique is mainly used with prediction models to evaluate how the designed model will perform on new unseen data. In predictive analysis, a model is generally trained on a dataset available with the required class labels, and when the model is built, it is used for testing on a new dataset with unknown output labels, so that model performance is evaluated (15). There are two techniques for dataset splitting: split data and cross-validation. K-fold cross-validation has an advantage over the split data technique because it works efficiently when applied to a biased dataset. K represents the number of times the data is trained with k-1 subsets for training and 1 subset for testing. In cross-validation, each instance is checked for both training and testing both. The ten-fold cross-validation is used to train and test the classification models.

RESULTS AND DISCUSSIONS

This section presents the results achieved and then briefly describes these results. The ANN classification algorithm applied

to feature selection using PSO. The performance of the classifiers is presented using standard measure of accuracy. Seven classification algorithms are applied to each dataset to compare their prediction performance with the proposed model. The prediction performance of ANN is improved using the PSO feature selection approach. Results with and without feature selection are compared to explore the impact of the most influential features and to present how the proposed feature optimization method performs better than other techniques. **Figure 3** presents the accuracy of the first breast cancer dataset and shows that logistic regression without feature selection (WFS) gives the highest accuracy of 93.15%, while the information gain increased the performance of ANN and achieved the highest accuracy of 93.67%. Finally, PSO is applied and a 5% increase in accuracy is observed. Therefore, the proposed model obtains the highest accuracy of 98.23%. For the second breast cancer dataset, as shown in **Figure 4**, Random Forest without feature selection gave the highest accuracy of 96.57%, followed by ANN with an accuracy of 96.43%, which increases to 97.43% after the selection of features by PSO. For diabetes prediction, as shown in **Figure 5**, ANN and Deep Learning achieve the highest accuracy of 78.72%. After feature extraction with PSO, the accuracy of ANN increases to 86.67%, which is the highest compared to other classification algorithms. When the proposed model is applied to another dataset for diabetes prediction, as shown in **Figure 6**, the highest accuracy of 93.59% is observed in ANN applied to PSO. Therefore, for both diabetes patient datasets, the proposed model reveals the highest results. For heart attack prediction, as shown in **Figure 7**, ANN and Deep Learning provide the highest accuracy of 82.17 and 82.13%, respectively. The most significant features are extracted with PSO, which shows an accuracy of 93.44% obtained by the proposed model. The k-NN shows a remarkable increase with an accuracy of 68.88% without feature selection and 90.62% with PSO. The proposed model is applied to another dataset for heart attack prediction, as shown in **Figure 8**, and PSO is found to show a remarkable increase in accuracy. ANN gave an accuracy of 81.85%, which increases to 85.56% after selecting the most significant features with PSO. **Figure 9** presents a comparison of eight classification algorithms with the proposed model for prediction of hepatitis disease. Random Forest achieved the highest accuracy of 95.13%, while ANN gave an accuracy of 93.46%. After feature selection, the proposed model obtained an accuracy of 98.46%. Next, the proposed model evaluates another dataset for hepatitis prediction in order to examine the results more accurately. **Figure 10** shows that the proposed model achieved highest accuracy of 73.73%. **Figure 11** shows the accuracy of kidney disease prediction. It shows that after PSO optimization, ANN and Random Forest give 98.90% accurate results. Information Gain shows a prominent increase in accuracy, but PSO significantly gives the best results. Different datasets have different accuracy because of different types of features and the number of instances. Optimization techniques help classifiers reduce irrelevant and useless features, remove outliers, and convert complex data into a simple form. It is found that overall, ANN, Random Forest, and Deep Learning give the best results, but ANN outperformed all other techniques. **Table 3** presents the highest accuracy gained by the proposed model for all five selected diseases. **Table 4** and **Figure 12** present the overall accuracy of the eight prediction algorithms achieved with PSO feature selection. An average of the results obtained by each prediction algorithm is calculated and it is found that the proposed ANN model produces the most accurate results.

It is shown that the proposed model gives the highest results, that is, 91.61%, followed by Random Forest with an accuracy of 91.19%. However, the Random Forest required more processing time compared to ANN. Therefore, the processing time of different classification techniques is also evaluated. Only one dataset for each disease is evaluated for processing time comparison. Total processing time is calculated in milliseconds (ms). **Figure 13** presents the total training and testing processing time for cancer, diabetes, heart, hepatitis, and kidney. The cancer dataset shows that Random Forest takes the most time. The prediction time required for the diabetes dataset also indicates that Random Forest requires more time compared to other classification techniques. The processing time for a heart attack shows that Random Forest followed by the SVM algorithm takes additional time. Hepatitis data processing time shows that SVM followed by Random Forest requires more time for training and testing. Random Forest and SVM are shown to be the slowest prediction and data processing algorithms for kidney disease. Overall, **Figure 13** shows that Random Forest and SVM required the most time for processing. Random Forest is an ensemble technique; its processing time depends on the number of trees combined to construct a forest, while SVM processing time depends on the kernel type. However, Naive Bayes and Logistic Regression take minimal time to build classification models. Furthermore, the processing time is also dependent on the number of instances in the dataset. Large datasets require more time for training and testing processes for all classification techniques. In (19), 10 different datasets are used for disease prediction with multiple feature selection approaches Information Gain, Gain Ratio, and ReliefF. Their proposed model gives the highest results of 99% when used with various subsets of patient attributes. Another study (75), suggested a feature selection method using Information Gain with different numbers of k for cross-validation to early predict kidney disease and achieved an accuracy of 81%. A hybrid classification model is proposed in (2) to diagnose three chronic diseases, including kidney disease, diabetes, and Hepatitis. The probabilistic feature selection approach is applied to a Logistic Regression classifier, and 91.6% accuracy is achieved. In another study (18), a chronic disease prediction model is built with three datasets of Breast Cancer, Diabetes and Heart attack patients. Convolutional Neural Network is applied to correlation coefficient-based attributes ranking technique. Diseases are predicted in two steps, at first with all available features and secondly with features of high significance. **Table 5** gives a comparative analysis of previous research studies as compared to the proposed approach.

CONCLUSION

Early diagnosis of chronic diseases is a major research challenge. Various artificial intelligence techniques are used in literature for medical data classification and disease prediction. Such techniques are often applicable for selected datasets to diagnose

specific disease with a limited set of attributes. In this paper, chronic diseases are predicted using an augmented artificial neural network based approach. The accuracy of the proposed model is improved using the particle swarm optimization (PSO) feature selection algorithm. This is shown to eliminate irrelevant features and assign weights to the most contributing features. Our proposed approach performs predictions more effectively and efficiently than other state-of-the-art models. Five most widespread chronic diseases are selected for diagnosis, including breast cancer, diabetes, heart attack, hepatitis, and kidney disease. Nine publicly available benchmark datasets are used in this study. The proposed model comprising an ANN classifier applied with PSO based feature selection is compared with seven state-of-the-art artificial intelligence techniques: decision tree, random forest, deep learning, naive Bayes, SVM, KNN, and logistic regression. Comparative results show that our proposed model performs the best out of all eight classification algorithms, with the highest accuracy of 99.67%. Our model obtained 98.23, 93.59, 98.46, 93.44, and 98.90% prediction accuracy for breast cancer, diabetes, hepatitis, heart, and kidney diseases. Whilst our proposed model produced the best results; the classification performance is found to depend significantly on the data features used for outcome prediction. Therefore, up to two datasets for each chronic disease are further used to apply our proposed model on different attributes. Results showed that our proposed approach outperformed other models. A comparison of processing time of classifiers showed that our optimized ANN approaches required less time compared to random forest, deep learning, and SVM. Our study can play a vital role for predicting multiple diseases using optimized ANN based prediction models in hospitals and medical labs. Further optimizing essential features can be beneficial for medical purposes to reduce patients' data attributes.

FUTURE WORK AND LIMITATIONS

In future, we plan to predict more chronic diseases with our proposed model. In particular, ensemble feature selection approaches may result in better prediction of multiple chronic diseases. Developing a system that can diagnose various diseases in patients is an open research problem and there has been little research work reported in this area. The number of patients living with more than one chronic disease is growing globally. We plan to build a prediction model based on classification and feature selection approaches to diagnose two or more chronic diseases in a person. A limitation of this study is that all classification methods are applied to selected diseases, while a dataset with different patient diseases may lead to suboptimal predictions. Although our augmented artificial intelligence based prediction model has shown promising results, there are a number of open research problems. First, generalized and globally applicable systems for disease prediction are not available. Second, prediction performance in research does not scale to real-world clinical applications due to limited data availability and different sets of attributes used in research studies. Third, real-time implementation and clinical validation of prediction models is a major challenge. Other open research areas include continuously improving prediction results based on clinical validation and diagnosing previously unknown diseases.

DATA AVAILABILITY STATEMENT

The publicly available datasets are used in this study, whose details are included in the "Materials and Methods" section of this article. Please contact the corresponding author for further requests.

AUTHOR CONTRIBUTIONS

JR and SB: conceptualization and methodology. JR, SB, JK, MW, AH, SJ, and RK: data curation. JR, JK, and AH: formal analysis, project administration, and writing—review and editing. JK and JR: funding acquisition. JR, SB, AH, SJ, and RK: investigation. MW, JR, and JK: resources. JR, JK, SJ, and RK: software. JR and SB: validation and writing—original draft. JR, SB, and JK: visualization. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. 2021R1A4A1031509) and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2020R1I1A3069700).

ACKNOWLEDGMENTS

Our deepest gratitude goes to the reviewers for improving the quality of the article.

REFERENCES

1. May HT, Anderson JL, Muhlestein JB, Knowlton KU, Horne BD. Intermountain chronic disease risk score (ICHRON) validation for prediction of incident chronic disease diagnoses in an australian primary prevention population. *Euro J Intern Med.* (2020) 79:81–87. doi: 10.1016/j.ejim.2020.06.009
2. Hegde S, Mundada MR. Early prediction of chronic disease using an efficient machine learning algorithm through adaptive probabilistic divergence based feature selection approach. *Int J Pervasive Comput Commun.* (2020) 20:145. doi: 10.1108/IJPCC-04-2020-0018
3. Latha CBC, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Inform. Med. Unlocked.* (2019) 16:1–9. doi: 10.1016/j.imu.2019.100203
4. Howard N, Chouikhi N, Adeel A, Dial K, Howard A, Hussain A. BrainOS: a novel artificial brain-alike automatic machine learning framework. *Front. Comput. Neurosci.* (2020) 14:1–15. doi: 10.3389/fncom.2020.00016

5. Bi X, Zhao X, Huang H, Chen D, Ma Y. Functional brain network classification for Alzheimer's disease detection with deep features and extreme learning machine. *Cognit Comput*. (2020) 12:513–27. doi: 10.1007/s12559-019-09688-2
6. Guo L. Under The background of healthy china: regulating the analysis of hybrid machine learning in sports activities to control chronic diseases. *Measurement*. (2020) 164:1–10. doi: 10.1016/j.measurement.2020.107847
7. W.H.O. *NonCommunicable Diseases*. (2018). Available online at: https://www.who.int/newsroom/fact-sheets/detail/noncommunicable-diseases (accessed December 12, 2021).
8. Hemanth Reddy K, Saranya G. "Prediction of cardiovascular diseases in diabetic patients using machine learning techniques," in *Artificial Intelligence Techniques for Advanced Computing Applications*, (New York, NY: Springer), p. 299–305 (2020).
9. W.H.O. *Cardiovascular diseases (CVDs)*. (2016). Available online at: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed December 12, 2021).
10. Diabetes - A Major Risk Factor for Kidney Disease. *National Kidney Foundation*. (2020). Available online at: https://www.kidney.org/atoz/content/diabetes (accessed December 12, 2021).
11. Parisi L, RaviChandran N. Evolutionary feature transformation to improve prognostic prediction of hepatitis. *Knowl Based Syst*. (2020) 200:1–10. doi: 10.1016/j.knosys.2020.106012
12. Abd El-Salam SM, Ezz MM, Hashem S, Elakel W, Salama R, ElMakhzangy H, et al. Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Inform. Med. Unlocked*. (2019) 17:1–7. doi: 10.1016/j.imu.2019.100267
13. Chugh G, Kumar S, Singh N. Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cogn Computat*. (2021) 13:1451–70. doi: 10.1007/s12559-020-09813-6
14. Raj RS, Sanjay DS, Kusuma M, Sampath S. "Comparison of support vector machine and naïve bayes classifiers for predicting diabetes," *In 2019* 1*st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*. (Mew York, NY: IEEE) (2019).
15. Aada A, Tiwari S. Predicting diabetes in medical datasets using machine learning techniques. *Int J Scientific Eng Res Vol.* (2017) 8:257–67. Available online at: https://ijsret.com/wp-content/uploads/2019/03/IJSRET_V5_issue2_154.pdf
16. Yuvaraj N, SriPreethaa KR. *Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster*. *Cluster Comput.* (2019) 22:1–9. doi: 10.1007/s10586-017-1532-x
17. Elhoseny M, Shankar K, Uthayakumar J. *Intelligent diagnostic prediction and classification system for chronic kidney disease*. *Sci Rep*. (2019) 9:1–14. doi: 10.1038/s41598-019-46074-2
18. Sandhiya S, Palani U. An effective disease prediction system using incremental feature selection and temporal convolutional neural network. *J Amb Intell Hum Comput*. (2020) 11:5547–60. doi: 10.1007/s12652-020-01910-6
19. Alam MZ, Rahman MS, Rahman MS. A Random Forest based predictor for medical data classification using feature ranking. *Inform Med Unlocked*. (2019) 15:1–12. doi: 10.1016/j.imu.2019.100180
20. Maini E, Venkateswarlu B, Marwaha D, Maini B. Upgrading the performance of machine learning based chronic disease prediction systems using stacked generalization technique. *Int J Comput Digit Syst.* (2020) 10:1–9. doi: 10.12785/ijcds/100192
21. Aldhyani TH, Alshebami AS, Alzahrani MY. Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms. *J Healthcare Eng*. (2020) 2020:1–16. doi: 10.1155/2020/4984967
22. Kanimozhi U, Ganapathy S, Manjula D, Kannan A. An intelligent risk prediction system for breast cancer using fuzzy temporal rules. *Nat Acad Sci Lett.* (2019) 42:227–32. doi: 10.1007/s40009-018-0732-0
23. Sinha A, Sahoo B, Rautaray SS, Pandey M, et al. "Improved framework for breast cancer prediction using frequent itemsets mining for attributes filtering," *In* 2019 I*nternational Conference on Intelligent Computing Control Systems (ICCS)*. (2019). (New York, NY: IEEE).
24. Liu N, Qi ES, Xu M, Gao B, Liu GQ. A novel intelligent classification model for breast cancer diagnosis. *Inf Process Manag.* (2019) 56:609–23. doi: 10.1016/j.ipm.2018.10.014
25. Islam M, Haque M, Iqbal H, Hasan M, Hasan M, Kabir MN. Breast cancer prediction: a comparative study using machine learning techniques. *SN Comput Sci.* (2020) 1:1–14. doi: 10.1007/s42979-020-00305-w
26. Alickovic E, Subasi A. Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Comput Appl.* (2017) 28:753–63. doi: 10.1007/s00521-015-2103-9
27. Azar AT, El-Said SA. Probabilistic neural network for breast cancer classification. *Neural Comput Applicat*. (2013) 23:1737–51. doi: 10.1007/s00521-012-1134-8
28. Alam TM, Iqbal MA, Ali Y, Wahab A, Ijaz S, Baig TI, et al. A model for early prediction of diabetes. *Inform Med Unlocked*. (2019) 16:100204. doi: 10.1016/j.imu.2019.100204
29. Oladele TO, Ogundokun RO, Kayode AA, Adegun AA, Adebiyi MO. "Application of data mining algorithms for feature selection and prediction of diabetic retinopathy," in *International Conference on Computational Science and Its Applications*. (New York, NY: Springer), (2019).
30. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Ra IH, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*. (2020) 9:274. doi: 10.3390/electronics9020274
31. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, Srivastava G. Deep neural networks to predict diabetic retinopathy. *J Amb Intell Human Comput*. (2020) 2020:1–14. doi: 10.1007/s12652-020-01963-7
32. Gadekallu TR, Iwendi C, Wei C, Xin Q. *Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning*. (2021).
33. Gadekallu TR, Pham QV, Huynh-The T, Bhattacharya S, Maddikunta PKR, Liyanage M. *Federated Learning for Big Data: A Survey on Opportunities, Applications, and Future Directions*. *arXiv preprint arXiv:2110.04160*. (2021)
34. Srinivasu PN, Bhoi AK, Jhaveri RH, Reddy GT, Bilal M. Probabilistic Deep Q Network for real-time path planning in censorious robotic procedures using force sensors. *J Real-Time Image Process.* (2021) 18:1773–85. doi: 10.1007/s11554-021-01122-x
35. Fitriyani NL, Syafrudin M, Alfian G, Rhee J. Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access*. (2019) 7:144777–89. doi: 10.1109/ACCESS.2019.2945129
36. Kanwal S, Rashid J, Kim J, Nisar MW, Hussain A, Batool S, et al. "An attribute weight estimation using particle swarm optimization and machine learning approaches for customer churn prediction," In: *2021 International Conference on Innovative Computing (ICIC)*. (2021) (Lahore: IEEE), p. 1–6.
37. Kanwal S, Rashid J, Nisar MW, Kim J, Hussain A. "An effective classification algorithm for heart disease prediction with genetic algorithm for feature selection," in *2021*. M*ohammad Ali Jinnah University International Conference on Computing (MAJICC)*. (2021) (Lahore: IEEE), p. 1–6.
38. Nazir T, Nawaz M, Rashid J, Mahum R, Masood M, Mehmood A, et al. Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model. *Sensors*. (2021) 21:5283. doi: 10.3390/s21165283
39. Rashid J, Shah SMA, Irtaza A. An efficient topic modeling approach for text mining and information retrieval through K-means clustering. *Mehran Univ Res J Eng Technol*. (2020) 39:213–22. doi: 10.22581/muet1982.2001.20
40. Rashid J, Shah SMA, Irtaza A. Fuzzy topic modeling approach for text mining over short text. *Inf Process Manag*. (2019) 56:102060. doi: 10.1016/j.ipm.2019.102060
41. Rashid J, Shah SMA, Irtaza A, Mahmood T, Nisar MW, Shafiq M, et al. Topic modeling technique for text mining over biomedical text corpora through hybrid inverse documents frequency and fuzzy k-means clustering. *IEEE Access*. (2019) 7:146070–80. doi: 10.1109/ACCESS.2019.2944973
42. Rashid J, Adnan Shah SM, Irtaza A. A novel fuzzy k-means latent semantic analysis (FKLSA) approach for topic modeling over medical and health text corpora. *J Intell Fuzzy Syst*. (2019) 37:6573–88. doi: 10.3233/JIFS-182776
43. Rashid J, Mahmood T, Nisar MW, Nazir T. "Phishing detection using machine learning technique," *In* 2020 First *International Conference of Smart Systems Emerging Technologies (SMARTTECH)*. (2020) (New York, NY: IEEE).
44. Ashraf R, Mehmood MS, Mahmood T, Rashid J, Nisar MW, Shah M. "An efficient forensic approach for copy-move forgery detection via discrete wavelet transform," *In 2020 International Conference on Cyber Warfare Security (ICCWS)*. (2020). (New York, NY: IEEE).

45. Fatima M, Nisar MW, Rashid J, Kim J, Kamran M, Hussain A. A novel fingerprinting technique for data storing and sharing through clouds. *Sensors.* (2021) 21:7647. doi: 10.3390/s21227647
46. Aksa M, Rashid J, Nisar MW, Mahmood T, Kwon HY, Hussain A. Bitmapaligner: bit-parallelism string matching with mapreduce and hadoop. *CMC-Comput Mater Continua.* (2021) 68:3931–46. doi: 10.32604/cmc.2021.016081
47. Iqbal H, Azhar E, Kanwal S, Rashid J, Nazir T. "Prevention of Beacons Collision in IEEE 802.15. 6 under coexistence," In *2020 International Symposium on Recent Advances in Electrical Engineering & Computer Sciences (RAEE & CS)*. (2020). (New York, NY: IEEE).
48. Zemmal N, Azizi N, Sellami M, Cheriguene S, Ziani A, AlDwairi M, et al. Particle swarm optimization based swarm intelligence for active learning improvement: Application on medical data classification. *Cogn Computat.* (2020) 12:991–1010. doi: 10.1007/s12559-020-09739-z
49. Upadhyay H, Juneja S, Juneja A, Dhiman G, Kautish S. Evaluation of ergonomics-related disorders in online education using fuzzy AHP. *Computation Intell Neurosci.* (2021) 21:4971. doi: 10.1155/2021/2214971
50. Juneja S, Juneja A, Dhiman G, Behl S, Kautish S. An approach for thoracic syndrome classification with convolutional neural networks. *Comput Math Methods Med.* (2021) 21:584. doi: 10.1155/2021/3900254
51. Juneja A, Juneja S, Soneja A, Jain S. Real time object detection using cnn based single shot detector model. *J Inform Technol Manage.* (2021) 13:62–80. doi: 10.22059/jitm.2021.80025
52. Juneja S, Jain S, Suneja A, Kaur G, Alharbi Y, Alferaidi A, et al. Gender and age classification enabled blockschain security mechanism for assisting mobile application. *IETE J Res.* (2021) 2021:1–13. doi: 10.1080/03772063.2021.1982418
53. Samant P, Agarwal R. Analysis of computational techniques for diabetes diagnosis using the combination of iris-based features and physiological parameters. *Neural Comput Applicat.* (2019) 31:8441–53. doi: 10.1007/s00521-019-04551-9
54. Wu JH, Li J, Wang J, Zhang L, Wang HD, Wang GL, et al. Risk prediction of type 2 diabetes in steel workers based on convolutional neural network. *Neural Comput Applicat.* (2019) 2019:1-16. doi: 10.1007/s00521-019-04489-y
55. Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access.* (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707
56. Khourdifi Y, Bahaj M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *Int J Intell Eng Syst.* (2019) 12:242–52. doi: 10.22266/ijies2019.0228.24
57. Rashid J, Nisar MW. How to improve a software quality assurance in software development-a survey. *Int J Comput Sci Inform Secur.* (2016) 14:99. Available online at: https://www.academia.edu/28701432/How_to_Improve_a_Software_Quality_Assurance_in_Software_Development-A_Survey
58. Kamran M, Rashid J, Nisar MW. Android fragmentation classification, causes, problems and solutions. *Int J Comput Sci Inform Secur.* (2016) 14:992. Available online at: https://www.academia.edu/30688611/Android_Fragmentation_Classification_Causes_Problems_and_Solutions
59. Khan M, Rashid J, Nisar MW. A CMMI complaint requirement development life cycle. *Int J Comput Sci Inform Secur.* (2016) 14:1000. Available online at: https://www.academia.edu/30688621/A_CMMI_Complaint_Requirement_Development_Life_Cycle
60. Rashid J, Nisar MW. A study on semantic searching, semantic search engines and technologies used for semantic search engines. *Int J Inform Technol Comput Sci (IJITCS).* (2016) 10:82–9. doi: 10.5815/ijitcs.2016.10.10
61. Adnan SM, Fatima T, Habiba U, Ilyas M, Adnan SM, Ahmad W, et al. Matching based algorithm for semantic search and its implementation. *Technic J.* (2018) 23:70–6. Available online at: https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/596/52
62. Tarawneh M, Embarak O. "Hybrid approach for heart disease prediction using data mining techniques," *In International Conference on Emerging Internetworking, Data & Web Technologies*. (2019) (New York, NY: Springer).
63. Reddy GT, Reddy M, Lakshmanna K, Rajput DS, Kaluri R, Srivastava G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol Intell.* (2020) 13:185–96. doi: 10.1007/s12065-019-00327-1
64. Batool S, Rashid J, Nisar MW, Kim J, Mahmood T, Hussain A. "A random forest students' performance prediction (rfspp) model based on students' demographic features," In: *(2021) Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*. New York, NY: Springer (2021).
65. Farooq K, Hussain A. A novel ontology and machine learning driven hybrid cardiovascular clinical prognosis as a complex adaptive clinical system. *Complex Adapt Syst Model.* (2016) 4:12. doi: 10.1186/s40294-016-0023-x
66. Parisi L, RaviChandran N, Manaog ML. A novel hybrid algorithm for aiding prediction of prognosis in patients with hepatitis. *Neural Comput Applicat.* (2020) 32:3839–52. doi: 10.1007/s00521-019-04050-x
67. Nilashi M, Ahmadi H, Shahmoradi L, Ibrahim O, Akbari E. A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *J Infect Public Health.* (2019) 12:13–20. doi: 10.1016/j.jiph.2018.09.009
68. Agarwal GG, Singh AK, Venkatesh V, Wal N. Determination of risk factors for hepatitis C by the method of random forest. *Annal Infect Dis Epidemiol.* (2019) 4:1. Available online at: http://www.remedypublications.com/open-access/determination-of-risk-factors-for-hepatitis-c-by-the-method-of-random-forest-1797.pdf
69. Nusinovici S, Tham YC, Yan MYC, Ting DSW Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *J Clin Epidemiol.* (2020) 122:56–69. doi: 10.1016/j.jclinepi.2020.03.002
70. Lu H, Uddin S, Hajati F, Moni MA, Khushi M. A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Appl Intell.* (2021) 2021:1–12. doi: 10.1007/s10489-021-02533-w
71. Damodara K, Thakur A. "Adaptive neuro fuzzy inference system based prediction of chronic kidney disease," *In 2021 7th International Conference on Advanced Computing Communication Systems (ICACCS)*. (2021) (New York, NY: IEEE).
72. Rady EHA, Anwar AS. Prediction of kidney disease stages using data mining algorithms. *Inform Me Unlocked.* (2019) 15:100178. doi: 10.1016/j.imu.2019.100178
73. Zhao J, Gu S, McDermaid A. Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression. *Math Biosci.* (2019) 310:24–30. doi: 10.1016/j.mbs.2019.02.001
74. Jongbo OA, Adetunmbi AO, Ogunrinde RB, Badeji-Ajisafe B. Development of an ensemble approach to chronic kidney disease diagnosis. *Scientific African.* (2020) 8:e00456. doi: 10.1016/j.sciaf.2020.e00456
75. Atallah DM, Badawy M, El-Sayed A, Ghoneim MA. Predicting kidney transplantation outcome based on hybrid feature selection and KNN classifier. *Multimed Tools Appl.* (2019) 78:20383–407. doi: 10.1007/s11042-019-7370-5
76. Lambert JR, Arulanthu P, Perumal E. "Identification of nominal attributes for intelligent classification of chronic kidney disease using optimization algorithm," *In. 2020 International Conference on Communication Signal Processing (ICCSP)*. (2020) (New York, NY: IEEE).
77. Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A. Multiple disease prediction using Machine learning algorithms. *Materials Today.* (2021) 7:361. doi: 10.1016/j.matpr.2021.07.361
78. *Breast Cancer Wisconsin (Diagnostic) Data Set*. Available Online At: https://www.kaggle.com/uciml/breast-cancer-wisconsin-data (accessed 26-10-2020).
79. *Breast Cancer Wisconsin*. Available Online At: https://github.com/zahangirbd/medical_data_for_classification/tree/master/data/Breast%20Cancer%20Wisconsin (accessed 26-10-2020).
80. *Pima Indians Diabetes Database*. Available Online at: https://www.kaggle.com/uciml/pima-indians-diabetes-database (accessed 26-10-2020).
81. *Diabetes Classification*. Available Online at: https://data.world/informatics-edu/diabetes-prediction (accessed 25-10-2020).
82. *Heart Disease UCI*. Available Online at: https://www.kaggle.com/ronitf/heart-disease-uci (accessed 25-10-2020).
83. *Heart Disease Prediction*. Available online at: https://data.world/informatics-edu/heart-disease-prediction (accessed 26-10-2020).
84. *Hepatitis*. Available online at: https://www.kaggle.com/harinir/hepatitis (accessed 26-10-2020).
85. *Indian Liver Patient Records*. Available online at: https://www.kaggle.com/uciml/indian-liver-patient-records (accessed 27-10-2020).

86. *Kidney Disease Dataset*. Available online at: https://www.kaggle.com/akshayksingh/kidney-disease-dataset (accessed 26-10-2020.)
87. Zhu C, Idemudia CU, Feng W. Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Inform Med Unlocked.* (2019) 17:100179. doi: 10.1016/j.imu.2019.100179
88. Joloudari JH, Saadatfar H, Dehzangi A, Shamshirband S. Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection. *Inform Med Unlocked*. (2019) 17:100255. doi: 10.1016/j.imu.2019.100255
89. Mujumdar A, Vaidehi V. Diabetes prediction using machine learning algorithms. *Procedia Comput Sci.* (2019) 165:292–9. doi: 10.1016/j.procs.2020.01.047
90. Usama M, Qadir J, Raza A, Arif H, Yau KLA, Elkhatib Y, et al. Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access.* (2019) 7:65579–615. doi: 10.1109/ACCESS.2019.2916648
91. Uddin S, Khan A, Hossain ME, Moni MA, et al. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* (2019) 19:1–16. doi: 10.1186/s12911-019-1004-8
92. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci.* (2019) 11:220. doi: 10.3389/fnagi.2019.00220
93. LaPierre N, Ju CJT, Zhou G, Wang W. MetaPheno: A critical evaluation of deep learning and machine learning in metagenome-based disease prediction. *Methods.* (2019) 166:74–82. doi: 10.1016/j.ymeth.2019.03.003
94. Mohammed AB. Decision tree, naïve bayes and support vector machine applying on social media usage in NYC/comparative analysis. *Tikrit J Pure Sci.* (2018) 22:94–9. Available online at: https://www.iasj.net/iasj/download/7fe80bcbbfee924d
95. Arunadevi DJ, Ramya S, Raja MR. *A* study of classification algorithms using Rapidminer. *Int J Pure Appl Math.* (2018) 119:15977–88. Available online at: https://www.acadpubl.eu/hub/2018-119-12/articles/6/1478.pdf

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Rashid, Batool, Kim, Wasif Nisar, Hussain, Juneja and Kushwaha. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Public Health

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fpubh.2022.829519

A Classification Algorithm-Based Hybrid Diabetes Prediction Model

Michael Onyema Edeh[1]*, *Osamah Ibrahim Khalaf*[2], *Carlos Andrés Tavera*[3], *Sofiane Tayeb*[4], *Samir Ghouali*[5,6], *Ghaida Muttashar Abdulsahib*[7], *Nneka Ernestina Richard-Nnabu*[8] and *AbdRahmane Louni*[9]

[1] Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria, [2] Al-Nahrain Nanorenewable Energy Research Center, Al-Nahrain University, Baghdad, Iraq, [3] COMBA R&D Laboratory, Faculty of Engineering, Universidad Santiago de Cali, Cali, Colombia, [4] Department of Electrical Engineering, Faculty of Sciences and Technology, Mustapha Stambouli University of Mascara, Mascara, Algeria, [5] Department of Electrical Engineering, Faculty of Science and Technology, University of Mascara, Mascara, Algeria, [6] STIC Laboratory, Faculty of Engineering Science, Tlemcen, Algeria, [7] Department of Computer Engineering, University of Technology, Baghdad, Iraq, [8] Department of Computer Science/Informatics, Alex Ekwueme Federal University Ndufu Alike Ikwo (AE-FUNAI), Abakaliki, Nigeria, [9] Faculty of Sciences and Technology, Mustapha Stambouli University of Mascara, Mascara, Algeria

Diabetes is considered to be one of the leading causes of death globally. If diabetes is not treated and detected early, it can lead to a variety of complications. The aim of this study was to develop a model that can accurately predict the likelihood of developing diabetes in patients with the greatest amount of precision. Classification algorithms are widely used in the medical field to classify data into different categories based on some criteria that are relatively restrictive to the individual classifier, Therefore, four machine learning classification algorithms, namely supervised learning algorithms (Random forest, SVM and Naïve Bayes, Decision Tree DT) and unsupervised learning algorithm (k-means), have been a technique that was utilized in this investigation to identify diabetes in its early stages. The experiments are per-formed on two databases, one extracted from the Frankfurt Hospital in Germany and the other from the database. PIMA Indian Diabetes (PIDD) provided by the UCI machine learning repository. The results obtained from the database extracted from Frankfurt Hospital, Germany, showed that the random forest algorithm outperformed with the highest accuracy of 97.6%, and the results obtained from the Pima Indian database showed that the SVM algorithm outperformed with the highest accuracy of 83.1% compared to other algorithms. The validity of these results is confirmed by the process of separating the data set into two parts: a training set and a test set, which is described below. The training set is used to develop the model's capabilities. The test set is used to put the model through its paces and determine its correctness.

Keywords: decision tree, random forest, Support Vector Machine (SVM), Bayesian Naive, diabetes, AI, ML, classification

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Abha Sharma,
GD Goenka University, India
Adesegun Osijirin,
Federal College of Dental Technology and Therapy Enugu, Nigeria
Praveen Kumar,
VIT University, India

***Correspondence:**
Michael Onyema Edeh
michael.edeh@ccu.edu.ng

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *05 December 2021*
Accepted: *03 February 2022*
Published: *31 March 2022*

Citation:
Edeh MO, Khalaf OI, Tavera CA, Tayeb S, Ghouali S, Abdulsahib GM, Richard-Nnabu NE and Louni A (2022) A Classification Algorithm-Based Hybrid Diabetes Prediction Model. Front. Public Health 10:829519. doi: 10.3389/fpubh.2022.829519

INTRODUCTION

Diabetes is a chronic disease also known as a silent disease. The World Health Organization (WHO) defines diabetes as a disease that prevents the body from properly using the energy provided by the food it consumes. In addition, the disease occurs when there are problems with the hormone insulin, which is naturally produced by the pancreas to help the body use sugar and fat and store some it (1).

More clearly, when we eat, food is broken down into glucose (sugar). This glucose provides energy for the body to function properly by drawing on its resources. During digestion, the blood carries the glucose throughout the body and supplies the cells. However, in order for the sugar in the blood to be delivered to the cells, the body needs insulin, a hormone secreted by the pancreas, which acts as a key to get the glucose from the blood into the cells of our body (2). There are three most common types of diabetes:

Type 1 diabetes: it is a condition in which the pancreas cannot produces enough insulin or does nor produces any insulin. It accounts for 5–10% of all diabetes cases, commonly affecting childhood and adolescence. Then, Type 2 diabetes–it is a condition in which the insulin produced does not effectively used to maintain the blood sugar level in the body. This type diabetes is common in people age 40 and above, but also appears in younger people. From all diagnosed diabetes cases worldwide, type 2 diabetes accounts for 90–95 percent. There is another type of diabetes called "gestational diabetes", caused by the lack response of the insulin receptors on the body tissues, even if the insulin levels are normal, which makes this condition different from the second type, and this case is very rare, account for 1–2% of all diabetes cased and it also increase the risk of developing type 2 diabetes later (3).

People with diabetes must be treated according to their type of diabetes. The goal of treatment is to keep the patient's blood sugar level within a normal range.

RELATED WORKS

Authors	Study abstract	Reference
Kumari and Chitra	In the proposed work, SVM with radial basis function kernel is used for classification. The performance parameters such as the classification accuracy (78.2 %), sensitivity (80%), and specificity of the SVM and RBF have found to be high thus making it a good option for the classification process.	(4)
Ahmed	Accuracy of the proposed models has been compared. The random forest method provided an accuracy of 74.7%, ANN gave 75.7% and K-means clustering method has given 73.6% accuracy.	(5)
Shetty et al.	Used K-Nearest Neighbors (KNN) and the Naïve Bayes technique for diabetes prediction. This technique was implemented in the form of a software program, in which users provide data in terms of patient records and the finding that the patient is diabetic or not.	(6)
Bhoia et al.	In this paper. Various supervised learning algorithms have been used such as CT, SVM, k-NN, NB, RF, NN, AB, and LR, and generated the training dataset and testing dataset using k-fold cross-validation with k = 10. The results of accuracy = 76.80%	(7)
Kandhasamy and Balamurali	The authors in used data from the University of California, diabetes mellitus patients were classified using a machine learning data repository to compare the performance of four common classifiers (J48 DT, the K-Nearest Neighbors algorithm, the Random Forest algorithm, and the Support Vector Machines algorithm). They used a data sample from the UCI machine learning data repository. Preliminary results suggest that the J48 DT classifier outperforms the other three classifiers in terms of accuracy (73.82 percent) before data preparation, and that the KNN (k = 1) and Random Forest classifiers outperform the other three classifiers after data pre-processing.	(8)
Vijayan et al.	The KNN method and the ANFIS algorithm are comparable. According to the results of the experiment, the amalgam of KNN and ANFIS gives the highest classification accuracy of 80 % among the algorithms tested.	(9)
Soleh et al.	The data in this study divided into two, 75% for training data, and 25% for testing data. This study produces an evaluation with an accuracy 80%, which means it is better than the previous paper, which is 75, 97%.	(10)
Rajput et al.	The target of analysis made in the present research is to list the risks factors and correlation that exist among those risk factors. In this work, logistic regression, support vector machine, random forest, decision tree, Naive Bayes, K nearest neighbor classifiers are used for prediction, and their accuracy is compared to choose the better machine learning model. SVM provides higher accuracy (96.0) among the choosen algorithms.	(11)
Deepa et al.	This work aims to propose an artificial intelligence-based intelligent system for earlier prediction of the disease using Ridge-Adaline Stochastic Gradient Descent Classifier (RASGD. The results of the proposed scheme have been compared with state-of-the-art machine learning algorithms such as support vector machine and logistic regression methods. The RASGD intelligent system attains an accuracy of 92%, which is better than the other selected classifiers.	(12)

Abbreviations: UCI, University California Irvine; DT, Decision Trees; PIDD, Pima Indians Diabetes Database; WHO, World Health Organization; ML, Machine Learning; AI, Artificial Intelligence; FBG, Fasting Blood Glucose; SVM, Support Vector Machine; SMO, Sequential Minimal Optimization; KNN, K-Nearest Neighbors.

METHODOLOGY USED

Model Diagram and Study Explanation

The proposed process is presented in the form of a model diagram in **Figure 1** below. The following diagram depicts the flow of the research done in the process of building the model:

In this study, we split the data set into two parts: a training set and a test set. The training set is used to train the model. The test set is used to test the model and evaluate the accuracy.

In the second step we use a K-means algorithm for data correction in order to improve performance and control the classified model (By changing the number of clusters).

Next, we invite the learned algorithms to be tested using the test database. Finally model evaluation with other related work.

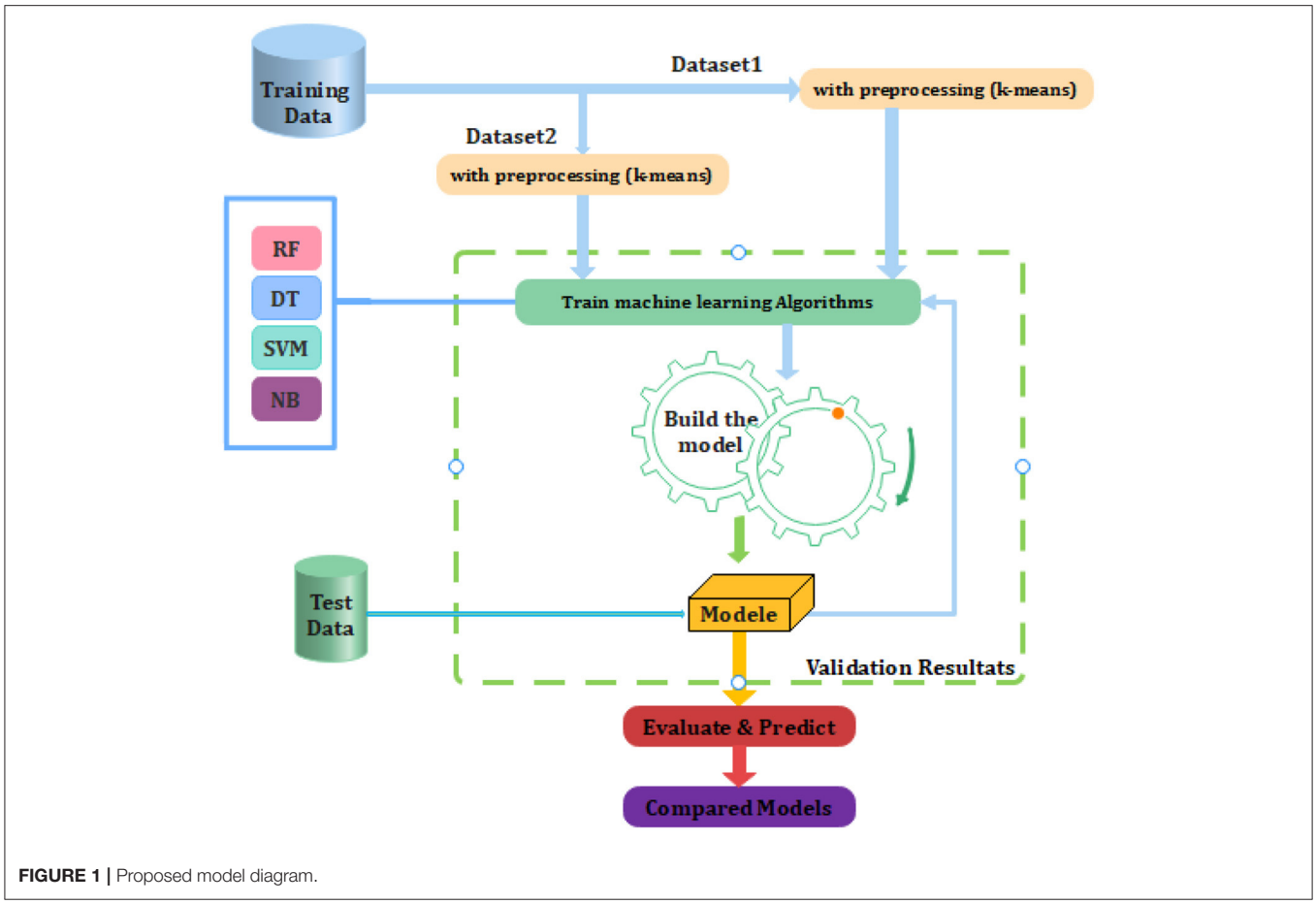


FIGURE 1 | Proposed model diagram.

Algorithms Used

Support Vector Machine

Support Vector Machine is based on statistical learning theory. SVMs were originally developed for binary classification, but can be effectively extended to multi-layer problems. SVM or Sequential Minimal Optimization (SMO) is a learning system that use a hypothesis space for linear functions in a high-dimensional space, and that has been trained using an optimization theory learning algorithm that employs a learning bias de-rived from statistical learning theory to achieve its results. SVM implements nonlinear class boundaries by translating nonlinear input vectors into a high-dimensional feature space using a linear model, which is implemented using the kernel of the SVM. Support vectors are training in-stances that are closer to the maximal hyper maximum level than the rest of the training examples. In order to define the binary layer boundaries support, all other training samples are rendered inapplicable.

As a result, the vectors are utilized to construct the ideal level hyper linear separation function (in the case of pattern recognition) or linear regression function (in the case of regression) in the feature space in question (13).

K-Means Clustering Algorithm

The k-means clustering algorithm is a machine learning algorithm that groups nearby points into clusters. In this algorithm, there is no learning model construction because we will locate the new point in any cluster based on its distance from all the clusters (mainly its distance from the cluster center or its arithmetic mean) and it is placed in the cluster that is closest. For example, imagine that you want to divide the points of a line into 3 groups. To Determine how close a point is to a particular group, we will use a measure of its distance from the group (for example, the distance between two points) (14).

Naive Bayes

NB is a classification approach in which the idea of independence and relatedness of all characteristics is defined as follows: Specifically, it specifies that the state of a given feature inside a class has no effect on the status of any other feature within the class. As a result of its foundation in conditional probability, it is regarded as a strong algorithm that may be utilized for classification applications. It performs effectively when dealing with data that has imbalance issues and missing values (15):

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \quad (1)$$

- P(A|B): conditional probability that the response variable has a certain value given the input characteristics. Additionally, this is referred to as the posterior probability.

- P(A): The response variable's a priori probability.
- P(B): The likelihood that the training data or evidence is correct.
- P(B|A): This is referred to as the probability training data.

Decision Tree

DT learning is one of the predictive modeling techniques used in statistics, data mining, and machine learning. Use a decision tree (as a predictive model) to move from observations on an item (represented in the branches) to conclusions about the target value of the item (represented in the paper). They use a hierarchical representation the data structure in the form of a sequence decisions (tests) to predict an outcome or a category. Each individual (or observation), which must be allocated to a class, is represented by a collection of variables, which are tested in the nodes of the tree. In the internal nodes, testing is carried out, and choices are taken in the paper nodes (16). In graph theory, a tree is a linked graph that is undirected, acyclic, and has no edges. There are three categories of nodes:

- Root node: This is the base of the tree and the most sensitive element when the tree is created and before it is exploded.
- Internal node: refers to nodes that have offspring that are themselves nodes.
- Final nodes: that do not contain any branches.

There are many DT's algorithms, and we can cite: ID3, C4.5, CART, C5, CHAID, SLIQ, UFFT, VFDT... (16).

Random Forest

Random Forest algorithm (17) is for statistics and machines that employs several learning methods to improve prediction performance. The two-part algorithm A. Tree bagging b. Each tree is produced from tree bagging to random forest:

1. Sample N instances at random - but with replacement, from the original data if the number cases are N inside the training set. The training set for developing the tree will be this sample.
2. A random number of characteristics and the optimal division utilized for dividing the node are picked when there are M input variables. During the forest growth, the value M is kept constant.
3. Each tree is cultivated as much as possible.

Dataset Used

We used two different databases in this study; Pima Indian Diabetes provided by the UCI Machine Learning repository (18) and a database extracted from the hospital in Frankfurt, Germany (19). Database extracted from the hospital in Frankfurt the first data is 2000 Pima Indian has 768 patient data with 8 attributes/features and one out-put with the patient's label/outcome (0: Not diabetic, 1: Diabetic). Two databases together consist some distinct medical variables, such as:

1. Pregnancies: number of pregnancies.
2. BMI: Body Mass Index (weight in kg / (height in m) ^{2}).
3. Insulin: Dose of insulin (mu U/ ml).
4. Age: Age at least 21 years.
5. Glucose: Plasma glucose concentration.
6. Blood Pressure: Diastolic blood pressure (mm Hg).
7. Skin Thickness: Thickness of the triceps skin fold (mm).
8. Diabetes Pedigree Function: Diabetes pedigree function (heredity).

And two classes (1 and 0)

- If class =1 implies diabetic patient.
- If class =0 implies non-diabetic patient.

The choice of these two bases is justified by the following criteria:

- The size of the database.
- Number of attributes.
- Number of classes.

Data Correction

Data cleaning is the next step in machine learning. It is considered one of the main steps in the working stages, and it is either building the model or breaking it. There is a saying that "the best data beats the most complex algorithms" in machine learning. Several aspects of data cleansing must be considered:

1. Discordances and omissions
2. Data mislabeling, same category repeated.
3. Invalid or missing data.
4. Outliers.

Unexpected Outliers in Both Bases

The observation and analysis of the two databases are presented in **Table 1**.

Methods of Handling Invalid Data Values

The Existing Works

- Ignore/delete these cases: delete all the observations with zero values but in this method we get a significant loss of data (about 50% of the data set).
- Put the mean values: calculate the median value of a specific column and replace this value in that column where we have zero.
- Avoid using parameters: The model can avoid using parameters with too many incorrect values. This may help thicken skin, although it's hard to tell.

In This Study

- Use of a classification algorithm: use a classification algorithm to recover the missing data where we have zero and replace them with the value found.

In our case, we have to apply this method and we have chosen the k-means algorithm with a variable K number of cluster (group) and replacing each column needs cleaning by the representative of cluster.

Evaluation Method

Train/Test

The data set is split into two parts: training and testing. The training set teaches the model. The test set is used to evaluate the model's correctness.

TABLE 1 | The observation and analysis of the two databases.

Variable	Observation and data analysis	Total number where the value is 0 (Pima base)	Total number where the value is 0 (Frankfort base)
Blood pressure	The data shows that there are 0 number for blood pressure.	35	90
glucose	It is impossible for a person to have a glucose zero value even if they are fasting.	5	13
Skin thikness	For normal people, the thickness of the skin fold cannot be <10 mm.	227	573
BMI	It is impossible for a person to have a BMI 0 value.	11	28
Insulin	In a rare situation a person may have 0 insulin	374	956
Pregnancies	It is normal to have a zero value for this column so there is no need for cleaning.	111	301

TABLE 2 | Accuracy measures.

Measures	Definitions	Formula
Accuracy	Accuracy determines the accuracy of the algorithm in predicting instances	A = (TP + TN)/ (Nombre total d'échantillons)
Recall	Is the ability of a classification model to identify all relevant instances	R = TP/(TP + FN)
F1– Mesure	Is the weighted average of precision and recall	F = 2 * [(P * R)/(P + R)]
Precision	Classifiers correctness/accuracy is measured by Precision	P = TP/ (TP + FP)

TABLE 3 | Evaluation attributes results for different models (Frankfurt Germany).

Algorithm used	Clusters	Accuracy	Recall score	F1 score
Naïve bayes	C = 16	0.776	0.625	0.654
SVM	C = 20	0.783	0.566	0.638
Decision tree	C = 285	0.971	0.975	0.958
Random forest	C = 20	0.989	0.95	0.972

TABLE 4 | Evaluation attributes results for the different models (Pima Indian).

Algorithm used	Clusters	Accuracy	Recall score	F1 score
Naïve bayes	C = 55	0.785	0.6	0.62
SVM	C = 16	0.831	0.533	0.648
Decision tree	C = 55	0.707	0.622	0.554
Random forest	C = 20	0.805	0.711	0.68

TABLE 5 | Comparison of the proposed work with the existing works (Pima Indian).

Method	Accuracy	Reference
Logistic regression	76.80%	(7)
Decision table	79.81%	(13)
Naïve Bayes	76.3%	(20)
Logistic Regression (LR)	80%	(10)
SVM	83.1%	Our study

Pima Indian Database

"Test size = 0.2." That is, 20% for the test and the rest 80% for the training.

Hospital Frankfurt Database

"Test size = 0.3." That is, 30% for the test and the rest 70% for the training.

Accuracy Measures

This study uses Naive Bayes, Random Forest, SVM, and DT algorithms. Train/Test Split is used in experiments. This study uses Accuracy, F1-Measure, Recall, and Precision metrics to classify. See **Table 2** for accuracy measures (20).

- True positives (TP), False positives (FP)
- True negatives (TN), False negatives (FN).

Experimental Results

This is clear from **Table 3**, which compares the different performance measures (accuracy, recall, and F1 score) used to evaluate the investigated machine learning models. Random Forest (RF) demonstrated the best accuracy when used in the tuned configuration (97.6 percent). In addition to Random Forest (RF), additional algorithms such as Decision Tree (DT) have demonstrated sufficient accuracy (97.5 percent). We can see that DT, Gaussian Naive Bayes, Random Forest, SVM, performed better. From the basic level, we can observe that Random Forest and DT work better than other algorithms (**Table 4**).

Based on different performance measures such as accuracy, recall, and F1 score, it is clear that the examined ML models are comparable; this is seen in the **Table 5**, SVM (Support Vector Machine) demonstrated the best accuracy when used in its optimized form (83.1 percent). Other algorithms, such as the Random Forest (RF), have demonstrated sufficient accuracy in addition to the Support Vector Machine (SVM) (80.5 percent). We can see that DT, Gaussian Naive Bayes, Random Forest, SVM, performed better. From the basic level, we can observe that Support Vector Machine and Random Forest work well than other algorithms.

From the results of this experimentation, we observe that the accuracy values for this database with all measurements are satisfactory with a disturbance sometimes the rate increases and

TABLE 6 | Comparison of the proposed work with the existing works (Frankfort Allemagne).

Algorithmes used	Data correction methods	Train/test split size	Accuracy	Reference
Random forest	The median	Test size = 0.2	91%	(21)
Gaussian process	The mean	Test size = 0.2	98.25%	(22)
DeepNN	Linear interpolation	Test size = 0.1	99.5%	(23–26)
Randomforest	k-means	Test size = 0.2	100%	Our Study
Randomforest	k-means	Test size = 0.3	97.6%	Our Study

sometimes it decreases by a small difference when changing the number of cluster as well as the random initialization of cluster centers can influence the results.

Note: After running the model several times, different results can be obtained in the same cluster number. This depends on the step of random initialization of the cluster centers.

According to the above table, the SVM model obtained the best accuracy which is equal to 83.1%. That is, among 153 attributes that were chosen for testing this model are classified 127 patients correctly. We select the SVM model as the most optimal model that works best for our dataset because it's high accuracy.

According to the above table, the Random Forest model obtained the best accuracy which is equal to 97.6%. That is, among 600 attributes that were chosen for testing this model are classified 582 patients correctly. We select the SVM model as the most optimal model that works best for our dataset because it's high accuracy.

DISCUSSION

Based on the results given in **Tables 5**, **6**, the goal of the four algorithms is to better classify future observations while reducing classification errors. it can be concluded that the suggested models are more accurate than other type 2 diabetes prediction models that have been investigated in the research indicated in these tables. When comparing the above findings, it is obvious that the notion of utilizing the k-means algorithm was successful in our work; as a consequence, we infer that improving the quality of the data enhances the outcomes. This is consistent with other studies (26–35) which shows the predictive accuracy of machine learning algorithms.

CONCLUSION

In this study, we proposed a supportive diagnosis system based on the comparison four models of prediction algorithms to predict diabetes in two different databases. On the basis of several performance assessment methodologies like as accuracy and recall, as well as the F1 score, different machine learning algorithms are compared and assessed. Using the classification results obtained, it can be concluded that the random forest machine learning technique provides more accurate prediction and higher performance than the other methods described in this study. However, when compared to other research accessible in the current literature, some of the other approaches utilized in this study, such as naive Bayes, DT and SVM, Random Forest, and others, produce the most optimum outcomes.

The main objective of this study is to help diabetologist to establish an accurate treatment routine for their diabetic patients. Due to the high accuracy and diagnose the disease in a shorter time and the rapid treatment, this study could open a window in the development of an electronic health system for diabetic patients. There are also a few aspects in this study that could be improved or expanded in the future. In perspective term:

- Creation of diabetes database for Algerian patients
- Diabetes prediction with the deep learning approach.
- Developed a solution based on an Android application in order to help people predict if they have diabetes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

ME: conceptualization, method, and analysis. OK: validation and conceptualization. ST: conceptualization and analysis. SG: methodology and investigation. NR-N: discussion and conclusion. CT: review, editing, and data curation. GA: results and conclusion. AL: editing and analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been funded by the Research General Direction at Universidad Santiago de Cali under call No 01-2021.

REFERENCES

1. Organisation mondiale de la santé. *Rapport mondial sur le diabète*. (2016). p. 88
2. Medtronic. Le Diabète En Quelque Mots. [En ligne]. Available online at: https://www.parlonsdiabete.com/parlonsdiabete/le-diabete-en-quelques-mots (accessed November, 2021).
3. Max Ray. *DIABETES -TYPE 2: The Review of Diabetic Studies*. (2019). Available online at: https://www.researchgate.net/publication/336634065_DIABETES_-TYPE_2 (accessed September, 2021).
4. Kumari VA, Chitra R. Classification de la maladie du diabète à l'aide d'une machine à vecteur de soutien. *IJERA*. (2013) 3:1797–801.
5. Ahmed TM. Using data mining to develop model for classifying diabetic patient control level based on historical medical records. *J Theor Appl Inf Technol*. (2016) 87:316–23.
6. Shetty D, Rit K, Shaikh S, Patil N. Diabetes disease prediction using data mining. In: *Innovations in information, embedded and communication systems (ICIIECS) international conference*. Coimbatore, India (2017). p. 1–5.
7. Bhoia SK, Pandab SK, Jenaa KK, Abhisekhc PA, Sahood KS, Samae NU, et al. 'Prediction of Diabetes in Females of PimaIndian Heritage: A

Complete Supervised Learning Approach'. *Turk J Comput Math Educ.* (2021) 12:3074–84.
8. Kandhasamy JP, Balamurali S. Performance analysis of classifier models to predict diabetes mellitus. *Procedia Comput Sci.* (2015) 47:45–51. doi: 10.1016/j.procs.2015.03.182
9. Vijayanv V, Ravikumar A. Study of data mining algorithms for prediction and diagnosis of diabetes mellitus. *Int J Comput Appl*. (2014) 95:12–6. doi: 10.5120/16685-6801
10. Soleh M, Ammar N, Sukmadi I. Website-based application for classification of diabetes using logistic regression method. *Jurnal Ilmiah Merpati.* (2021) 9. Retrieved from: https://ojs.unud.ac.id/index.php/merpati/article/view/66691 (accessed November, 2021).
11. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmanna K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Humaniz Comput*. (2021).
12. Deepa N, Prabadevi B, Maddikunta PK, Gadekallu TR, Baker T, Khan MA, et al. An AI-based intelligent system for healthcare analysis using Ridge-Adaline Stochastic Gradient Descent Classifier. *J Supercomput*. (2021) 77:4–16. doi: 10.1007/s11227-020-03347-2
13. Kumar DA, Govindasamy R. Performance et évaluation des techniques d'exploration de données de classification dans le diabète. *IJCSIT*. (2015) 6:1312–9.
14. Ouamri O. Contribution des arbres dirigés et les k-means pour l'indexation et recherche d'images par contenu, Mémoire de Magister en Informatique, by H. fizazi izabatene, Université des Sciences et de la Technologie d'Oran-Mohamed Boudiaf, Département d'Informatique (2011). p. 113
15. Java T point. Naïve Bayes Algorithm. [En ligne]. Available online at: https://www.javatpoint.com/machinelearning-naive-bayes-classifier (accessed July, 2021).
16. Rokach L, Maimon. *Exploration de données avec arbres de décision: théorie et applications*. 2e édition, World Scientific Pub Co Inc. (2015). Available from: http://cedric.cnam.fr/vertigo/cours/ml2/coursArbresDecision.html
17. Kandhasamy JP, Balamurali S. *Procedia Computer Science.* Elsevier (2015). Available online at: https://scholar.google.com/citations?user=LYSOeWMAAAAJ&hl=fr&oi=sra
18. Kaggle, UCI Machine Learning, Base de donnée Pima Indian Diabetes. Available online at: https://www.kaggle.com/uciml/pima-indians-diabetes-database (accessed November, 2021).
19. Kaggle, Johan Ensemble de données sur le diabète, extrait de l'hôpital de Francfort, Allemagne. Available online at: https://www.kaggle.com/johndasilva/diabetes (accessed November, 2021).
20. Sisodia D, Sisodia DS. Prediction of Diabetes using Classification Algorithms. *Procedia Comput Sci.* (2018) 132:1578–85. doi: 10.1016/j.procs.2018.05.122
21. Amel S, Karima R. The prediction of diabetes using machine learning algorithms, Master's thesis in Computer Science, supervised by Brahimi Farida, University AMO of Bouira Faculty of Sciences and Applied Sciences, Department of Computer Science (2019). p. 85.
22. Nishat MM, Faisal F, Mahbub MA, Mahbub MH, Islam S, Hoque MA. Performance assessment of different machine learning algorithms in predicting diabetes mellitus. *Biosc Biotech Res Comm*. (2021) 14. doi: 10.21786/bbrc/14.1/10
23. Daanouni O, Cherradi B, Tmiri A. Predicting Diabetes Diseases Using Mixed Data and Supervised Machine Learning Algorithms. SCA2019 (2019).
24. Onyema EM. Opportunities and challenges of use of mobile phone technology in teaching and learning in Nigeria-a review. *IJREI*. (2019) 3:352–8. doi: 10.36037/IJREI.2019.3601
25. Onyema EM, Elhaj MAE, Bashir SG, Abdullahi I, Hauwa AA, Hayatu AS. Evaluation of the Performance of K-Nearest Neighbor Algorithm in Determining Student Learning Styles. *Int J of Innovative Sci, Eng & Techn.* (2020) 7:91–102.
26. Shariq AB, Muhammad WA, Syed AH, Arindam G, Onyema EM. Smart Health Application for Remote Tracking of Ambulatory Patients. In: Hafizul Islam SK, Samanta D, editors. *Smart Healthcare System design: Security and Privacy Aspects*. Wiley Online Library (2021). p. 33–55.
27. Jo O, Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
28. Celestine I, Suresh P. An efficient and unique TF/IDF algorithmic model-based data analysis for handling applications with big data streaming. *Electronics*. (2019) 8:28. doi: 10.3390/electronics8111331
29. Celestine I, Mueen U, James A, Pascal N, Joseph HA, Ali KB. On detection of Sybil attack in large scale VANETs using spider-monkey technique. *IEEE Access.* (2018) 6:47258–67. doi: 10.1109/ACCESS.2018.2864111
30. Jalil P, Celestine I, Praveen K, Gadekallu TR, Lakshmanna K, Bashir AK. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw Pract Exp*. (2020) 14:2–9. doi: 10.1002/spe.2797
31. Rajendran S, Khalaf OI, Alotaibi Y, Alghamdi S. MapReduce-based big data classification model using feature subset selection and hyperparameter tuned deep belief network. *Sci Rep*. (2021) 11:24138. doi: 10.1038/s41598-021-03019-y
32. Rajalakshmi M, Saravanan V, Arunprasad V, Romero CAT, Khalaf OI, Karthik C. Machine learning for modeling and control of industrial clarifier process. *IASC.* (2022) 32:339–59. doi: 10.32604/iasc.2022.021696
33. Alsubari SN, Deshmukh SN, Alqarni AA, Alsharif N, Aldhyani THH, Alsaade FW, et al. Data Analytics for the Identification of Fake Reviews Using Supervised Learning. *Comput Mater Contin*. (2022) 70:3189–204. doi: 10.32604/cmc.2022.019625
34. Khalaf OI, Abdulsahib GM. Optimized dynamic storage of data (ODSD) in IoT based on blockchain for wireless sensor networks. *Peer-to-Peer Netw Appl*. (2021) 14:2858–73. doi: 10.1007/s12083-021-01115-4
35. Surendran R, Khalaf OI, Andres C. Deep learning based intelligent industrial fault diagnosis model. *Comput Mater Contin.* (2022) 70:6323–38. doi: 10.32604/cmc.2022.021716

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Edeh, Khalaf, Tavera, Tayeb, Ghouali, Abdulsahib, Richard-Nnabu and Louni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpubh.2022.860381

Identification of a Four-Gene Metabolic Signature to Evaluate the Prognosis of Colon Adenocarcinoma Patients

Yang Zheng[1,2†], *Rilige Wu*[3†], *Ximo Wang*[1,2,4*] *and Chengliang Yin*[5*]

[1] Graduate School, Tianjin Medical University, Tianjin, China, [2] Tianjin Key Laboratory of Acute Abdomen Disease Associated Organ Injury and ITCWM Repair, Institute of Integrative Medicine for Acute Abdominal Diseases, Tianjin Nankai Hospital, Tianjin, China, [3] College of Science, Beijing University of Posts and Telecommunications, Beijing, China, [4] Tianjin Haihe Hospital, Tianjin, China, [5] Faculty of Medicine, Macau University of Science and Technology, Macau, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Guohui Zhang,
Capital Medical University, China
Kun Liu,
Sir Run Run Shaw Hospital, China
Praveen Kumar,
VIT University, India
Weizheng Wang,
City University of Hong Kong, Hong Kong SAR, China

***Correspondence:**
Chengliang Yin
chengliangyin@163.com
Ximo Wang
wangximonkyy@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 22 January 2022
Accepted: 14 March 2022
Published: 07 April 2022

Citation:
Zheng Y, Wu R, Wang X and Yin C (2022) Identification of a Four-Gene Metabolic Signature to Evaluate the Prognosis of Colon Adenocarcinoma Patients. Front. Public Health 10:860381. doi: 10.3389/fpubh.2022.860381

Background: Colon adenocarcinoma (COAD) is a highly heterogeneous disease, thus making prognostic predictions uniquely challenging. Metabolic reprogramming is emerging as a novel cancer hallmark that may serve as the basis for more effective prognosis strategies.

Methods: The mRNA expression profiles and relevant clinical information of COAD patients were downloaded from public resources. The least absolute shrinkage and selection operator (LASSO) Cox regression model was exploited to establish a prognostic model, which was performed to gain risk scores for multiple genes in The Cancer Genome Atlas (TCGA) COAD patients and validated in GSE39582 cohort. A forest plot and nomogram were constructed to visualize the data. The clinical nomogram was calibrated using a calibration curve coupled with decision curve analysis (DCA). The association between the model genes' expression and six types of infiltrating immunocytes was evaluated. Apoptosis, cell cycle assays and cell transfection experiments were performed.

Results: Univariate Cox regression analysis results indicated that ten differentially expressed genes (DEGs) were related with disease-free survival (DFS) (P-value$<$ 0.01). A four-gene signature was developed to classify patients into high- and low-risk groups. And patients with high-risk exhibited obviously lower DFS in the training and validation cohorts (P $<$ 0.05). The risk score was an independent parameter of the multivariate Cox regression analyses of DFS in the training cohort (HR $>$ 1, P-value$<$ 0.001). The same findings for overall survival (OS) were obtained GO enrichment analysis revealed several metabolic pathways with significant DEGs enrichment, G1/S transition of mitotic cell cycle, CD8+ T-cells and B-cells may be significantly associated with COAD in DFS and OS. These findings demonstrate that si-FUT1 inhibited cell migration and facilitated apoptosis in COAD.

Conclusion: This research reveals that a novel metabolic gene signature could be used to evaluate the prognosis of COAD, and targeting metabolic pathways may serve as a therapeutic alternative.

Keywords: COAD, metabolic, LASSO, Cox, signature, prognostic prediction

INTRODUCTION

Metabolic reprogramming has been recognized as one of the distinguishing feature of tumor cells recently (1). Cancer metabolism is a major contributor to tumor initiation, growth, and metastasis in colorectal cancer (2).

Colon adenocarcinoma (COAD) is a primary intestinal aggressive malignancy (3). It is the third most prevalent cancer (10.0%) and is ranked second (9.4%) in cancer-related mortality worldwide (4). Further, this disease is genetically heterogeneous whose prognosis is uniquely challenging. China, with the most population in the world, is also ranked first in terms of new cases of cancer and cancer-related deaths in the world (5, 6). In China, more than 555,000 new colon and rectum cancer cases are projected to be discovered annually (4, 6). The world's cancer burden including COAD continues to increase and malignancy is quickly becoming the foremost cause of human death in this century (7). Therefore, predicting the prognosis of COAD patients and identifying new therapeutic strategies for COAD are critical.

According to previous reports, metabolism serves a vital role in the progress of COAD and some key genes such as OMA1 mediate metabolism reprogramming under hypoxia, thus facilitating colorectal cancer development by promoting the Warburg effect (8). And LINRIS stabilizes IGF2BP2 and facilitates aerobic glycolysis in COAD (9). However, the link between metabolic-related genes and the prognosis of COAD patients remains largely uncharacterized. Metabolism and expression of specific genes affect the occurrence of metabolism, making it a therapeutic potential target for the management of malignant tumor (10). Thus, it is important to gain further understanding of the underlying metabolic mechanisms of COAD tumorigenesis and progression. Further, it is needed urgently to identify new prognostic-related markers which could become therapeutic potential targets in COAD (11, 12).

First, we acquired expression profiles of mRNA and the relevant clinical features of COAD patients from open data resources in our research. We then established a multigene prognostic characteristic of COAD by metabolic-related differentially expressed genes (DEGs) in a Gene Expression Omnibus (GEO) cohort. Finally, functional enrichment analysis was performed to explain the potential mechanisms involved, after which a prognostic model was developed based on several metabolic-related genes. We consider that this prognostic model with the strong predictive power of COAD will improve prognosis risk assessment of patients with COAD and help to develop a more accurate evaluation of their clinical management.

MATERIALS AND METHODS

Publicly Available mRNA Data and Metastasis Gene Sets

An RNA-sequencing (RNA-seq) dataset and relevant clinical data of COAD patients were acquired from The Cancer Genome Atlas colon adenocarcinoma dataset (TCGA-COAD) and GEO website (GSE25071) to compare human colorectal tumors and normal colorectal tissues. The normalization of gene expression profiles was performed by the scale method provided in the "limma" package in R. We then identified DEGs though the comparison of patients with COAD with recurrent and normal humans in the GEO cohort (GSE21510 and GSE32323). A total of 566 COAD patients from the training set of the TCGA-COAD cohort and 573 COAD patients from the GSE39582 cohort used for the validation set were eventually enrolled. A list of metabolic-related genes was then gathered from the Gene Set Enrichment Analysis (GSEA, v3.0, http://software.broadinstitute.org/gsea/index.jsp) website (13) and is exhibited in **Supplementary Table 1**.

Building and Verification of a Prognostic Model of Metabolic-Related Genes

DEGs with log (|fold change|) >1 and a false discovery rate (FDR) < 0.05 between tumor tissues and adjacent non-cancerous tissues in the TCGA-COAD and GSE25071 datasets were identified by the "limma" package in R. Differences between the recurrence and normal human GEO cohorts (GSE21510 and GSE32323, respectively) were also evaluated. Univariate Cox analysis of disease-free survival (DFS) and overall survival (OS) in DEGs were carried out to identify the metabolic-related genes with prognostic significance. We utilized the log-rank test to adjust the *p*-value.

To reduce the number of genes, we used the least absolute shrinkage and selection operator (LASSO) Cox regression analysis to screen genes for DFS and OS, respectively. The input of the LASSO Cox regression analysis were the above-mentioned screened genes. The most vital value in the LASSO Cox regression analysis was λ. Using different λ values, LASSO Cox regression analysis can be utilized to screen for different genes. We used cross-validation to select the optimal λ for DFS and OS. With these λ values, the genes associated with DFS and OS were identified via LASSO Cox regression analysis, respectively. These genes were then subjected to multivariate Cox regression analysis to determine their coefficients for the prognostic models of DFS and OS, respectively. The prognostic models were represented as a risk score, which was expressed as follows:

$$\textit{Risk Score} = \sum_{i=1}^{N} \textit{Expi} \times \textit{Coefi}$$

A total of 222 patients obtained from the TCGA-COAD cohort with a DFS survival status and survival time were classified into low- and high-risk groups by the best cut-off point determined by the surv_cutpoint function in the "Survminer" package in R. We then plotted the survival curves for the risk score via the "survfit" function with the "survival" R package. We did the same for the 214 patients included in the TCGA-COAD data set who had the stages (stages I, II, and III) information of all those 222 patients, 51 patients for stage I, 147 patients for stage I and II, and 67 patients for stage III. To appraise the predictive power of the risk score, we plotted the receiver operating characteristic (ROC) and displayed the 1-, 3- and 5-year projections as function of survival ROC in the "survivalROC" R package.

Validation of the Prognostic Model as an Independent Clinical Factor

To explore the correlation of the established prognostic model with the clinical information, we carried out univariate cox regression analyses to generate risk scores and consider other clinical features including stage, sex, and age in 214 DFS patients included in the TCGA-COAD data set. Risk scores and other clinical factors were deemed statistically relevant at a *p*-value threshold of <0.05. Association between risk score levels and other clinical factors with DFS and OS were investigated by multivariate Cox proportional hazard models. The results of the multivariate Cox regression model were visualized in a combined forest plot and the same procedure was also performed in TCGA OS patients and the validation data set.

Verification of the Nomogram

A nomogram was created to forecast the DFS of patients at 1-, 3-, and 5-year survival probabilities to help doctors understand and apply the model. Then, we performed ROC analysis to gain the area under the curve (AUC) and verify the predictive effect of the model. The same procedure was performed in TCGA OS patients and the validation data set. The calibration curve of the nomogram was drawn to validate the nomogram's predictive value. The calibration curve was produced by using the calibration function in the "rms" package in R.

The decision curve analysis (DCA) results were plotted to quantify and assess the clinical value of the nomogram (14). DCA can be performed to obtain the clinical net benefit of the nomogram compared with all or none of the strategies (15).

Functional Enrichment Analysis and Cell Immune Infiltration

A comprehensive knowledge-base update to the sixth version of the original web-based programs is provided by the Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.8, https://david.ncifcrf.gov/). With the comprehensive set of function annotation tools provided by DAVID, researchers can examine the biological meanings of a large set of genes. The Gene Ontology resource (GO; http://geneontology.org) provides structured, computable knowledge regarding the functions of genes and gene products (16). The knowledge of GO resource is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research. DAVID was used to conduct GO analysis based on genes in a univariate Cox analysis of OS and DFS.

Tumor Immune Estimation Resource (TIMER; https://cistrome.shinyapps.io/timer/) is a web interactive platform and is used to analyse tumor-infiltrating immunocytes systematically. We utilized TIMER to study the correlation of genes in the risk score model and the signatures of tumor-infiltrating immunocytes in COAD. The "Gene" module was carried out to investigate the relevance between the risk score model expression and immunocyte infiltration levels specific to each gene (B-cells, neutrophils, macrophages, dendritic cells, CD4+ and CD8+ T-cells) with the TCGA database. We applied TIMER to explore the relationship between the gene expression in the risk score model and the marker gene sets of different immunocytes with the "Correlation" module. The relationship between the gene expression in the risk score model and the tumor-infiltrating immune cells were assessed by purity-correlated partial Spearman's correlation and statistical significance (17).

External Validation Using Online Databases

Hub genes in LASSO Cox were surveyed from the following online databases: (1) Oncomine database analysis. Oncomine database (https://www.oncomine.org/resource/main.html) is a tumor microarray database and online data analysis tool that collects many "multi-arrays" (18). This tool was used to determine gene expression signatures. Gene expression levels in various types of cancers were identified using the Oncomine database at a *p*-value threshold of 0.001 and a fold change threshold of 2 using gene ranking. The following conditions were used to acquire the mRNA expression level in tumor tissue compared with normal tissues: $P < 1E\text{-}4$, fold change > 2, and top gene rank 10% (19). (2) cBioPortal analysis. The cBioPortal for Cancer Genomics (http://cbioportal.org) was constructed specifically to decrease the difficulty of obtaining complex data sets and promote the translation of genomic data into novel biological knowledge, treatments, and clinical trials (20). This platform encourages the study of multidimensional tumor genomics data by supporting visualization and estimating across genes, samples, and data types. The resource allows researchers to visualize the patterns of gene alterations across samples in one tumor research, to compare the frequencies of gene alterations across multiple tumor researches, and to aggregate alterations of all related genomes in an single cancer sample. cBioPortal encompasses multiple genomic data types such as somatic mutations, DNA copy number alterations (CNAs), mRNA and microRNA (miRNA) expression, DNA methylation, protein abundance, and phosphoprotein abundance (21). (3) PrognoScan database Analysis. The PrognoScan database (http://www.abren.net/PrognoScan/) was utilized to explore the association between gene expression and survival across various types of cancers. PrognoScan was utilized to dissect the correlations between gene expression and prognosis indicators (e.g., OS and DFS) across a large number of open tumor microarray datasets. The threshold of Cox *p*-value was adjusted to <0.05 (22). (4) GEPIA database Analysis. The Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/) database is a developed interactive web server for analyzing the RNA sequencing expression data of tumor and normal samples from the TCGA and the GTEx projects, using a standard processing pipeline. GEPIA provides customizable functions such as tumor/normal differential expression analysis, profiling according to cancer types or pathological stages, patient survival analysis, similar gene detection, correlation analysis and dimensionality reduction analysis (23).

Apoptosis and Cycle Assay

The analysis of the previous sections indicated that FUT1 was an aberrantly expressed gene in COAD tumors and therefore high FUT1 expression could be used as a predictor of adverse

TABLE 1 | Clinical characteristics of colorectal patients.

Characteristics	TCGA (*N* = 566)	GSE39582 (*N* = 573)
AGE, Median (Q1,Q3)	67 (57, 75)	68 (59, 76)
Gender, *n* (%)		
Female	264 (47%)	256 (45%)
Male	302 (53%)	317 (55%)
ANNARBOR.Stage, *n* (%)		
NA	12 (2%)	–
0	–	4 (1%)
I	101 (18%)	37 (6%)
II	207 (37%)	265 (46%)
III	165 (29%)	208 (36%)
IV	81 (14%)	59 (10%)

outcomes. So we examined the function of FUT1 in cell apoptosis and cell cycle and performed a flow cytometry assay. We cultured HCT-116 cells for 24 h with FUT1 and then harvested the cells. Next, Annexin V and propidium iodide (PI) were used in accordance with the manufacturer's recommendation (Beijing Solarbio, China). The apoptotic rate and cell cycle were detected using a NovoCyte flow cytometer (ACEA, Biosciences, USA) and analyzed using the NovoExpress software (ACEA, Biosciences, USA).

Cell Transfection

Cell transfection was also performed to further verify the effect of FUT1. HCT-116 cells were gained from the American Type Culture Collection. The cells were kept in Dulbecco's modified Eagle's medium (DMEM) (Gibco, USA), supplemented with 10% fetal bovine serum (FBS) (Gibco, USA), 100 μg/mL streptomycin (Gibco, USA) and 100 U/mL penicillin (Gibco, USA) in a humidified incubator with 37°C and 5% CO_2. FUT1 was silenced in HCT-116 cells by transfection with FUT1 short interfering RNA (si-FUT1). si-FUT1 and the siRNA negative control (si-NC) were purchased from Tsingke Biotech (Beijing, China). Lipofectamine 3000 (Invitrogen, Carlsbad, CA, USA) was used to carried out the transfection trials according to the manufacturer's instructions.

RESULTS

Patient Selection

A total of 566 COAD patients were selected as the training set from TCGA and 573 COAD patients were selected from the GSE39582 cohort as the validation set. The specific clinical features of the patients were listed in **Table 1**. A flow chart of our research was illustrated in **Supplementary Figure 1**.

Identifying of Prognostic Metabolic-Related DEGs in the TCGA Cohort

A total of 153 overlapping DEGs were identified between (1) tumor tissues and non-tumor adjacent tissues in TCGA and GSE20571, and (2) recurrent and normal human tissues in the GEO cohort (GSE21510 and GSE32323). A total of 62 genes were identified by matching 153 DEGs with metabolic-related genes, 21 of which were correlated with OS and 10 were correlated with DFS in the univariate Cox regression analysis ($P < 0.01$) in the training set (**Tables 2**, **3**).

TABLE 2 | Disease-free survival associated gene list in COAD from the training set (TCGA).

No.	Gene	HR	*P*-value
1	CCND1	0.238875	0.000367
2	CXCL1	2.968093	0.007837
3	CXCL10	3.176181	0.002312
4	CXCL3	2.856345	0.005621
5	EDAR	0.190756	3.90E-05
6	FUT1	0.358673	0.008883
7	INHBA	0.31451	0.001944
8	MMP1	4.725989	0.000239
9	MYC	2.978401	0.00377
10	PPAT	2.960687	0.006917

TABLE 3 | Overall survival associated gene list in COAD from the training set (TCGA).

No.	Gene	HR	*P*-value
1	CCNB1	2.639522	0.000693
2	CCND1	0.535604	0.005232
3	CDC6	1.75829	0.002691
4	CDH3	1.924981	0.008129
5	CXCL1	1.743645	0.005178
6	CXCL11	1.937543	0.001367
7	CXCL3	1.765604	0.00236
8	FUT1	0.583245	0.007906
9	GPD1L	1.832985	0.004354
10	MAD2L1	1.977956	0.000292
11	MMP1	2.060513	0.000426
12	MYC	1.670215	0.009585
13	NME1	2.070147	0.001144
14	PGM1	2.00308	0.001948
15	PPAT	2.180598	0.001574
16	PPIL1	1.822861	0.001349
17	PTTG1	1.746729	0.003743
18	RIPK2	2.426172	0.000253
19	RRM2	1.896863	0.001508
20	SPP1	0.519043	0.000739
21	TKT	1.957591	0.000346

Establishment of a Prognostic Model in the TCGA Cohort

We used LASSO Cox regression analysis to both the 10 DFS genes and 21 OS genes to develop prognostic models with the expression profile of the above-mentioned genes. LASSO Cox

regression analysis was used to screen four genes (CCND1, EDAR, FUT1, and PPAT) in DFS and eight genes (CCNB1, CDC6, FUT1, GPD1L, MAD2L1, MMP1, SPP1, and TKT) in OS, which acted as the input for multivariate Cox regression analysis. We obtained the coefficients of the prognostic models of both DFS and OS using multivariate Cox regression analysis (**Figures 1A,B,D,E**). Finally, prognostic models were displayed in the form of a risk score as follows:

$$\begin{aligned} \textit{Risk score of DFS} &= \mathbf{0.411228099154717} \times \textit{CCND1} \\ &+ \mathbf{0.547206032433896} \times \textit{EDAR} \\ &+ \mathbf{0.302185127495718} \times \textit{FUT1} \\ &- \mathbf{0.583673054198275} \times \textit{PPAT} \\ \textit{Risk score of OS} &= -\mathbf{0.0985300797882395} \times \textit{CCNB1} \\ &- \mathbf{0.0566241392969015} \times \textit{CDC6} \\ &+ \mathbf{0.111288163590287} \times \textit{FUT1} \\ &- \mathbf{0.144434028275748} \times \textit{GPD1L} \\ &- \mathbf{0.21174022083491} \times \textit{MAD2L1} \\ &- \mathbf{0.231299794085908} \times \textit{MMP1} \\ &+ \mathbf{0.129051446614179} \times \textit{SPP1} \\ &- \mathbf{0.140224178453846} \times \textit{TKT} \end{aligned}$$

GO term analyses were carried out to investigate the potential biological meanings of the 21 identified DFS genes and 10 OS genes. As depicted in **Figures 1C,F**, GO annotation revealed top categories that were positively correlated with metabolism in DFS such as chemokine-mediated signaling pathway, G1/S transition of mitotic cell cycle, and re-entry into mitotic cell cycle, among others. Other pathways were identified for OS including G1/S transition of mitotic cell cycle, cell division, and inflammatory response. These results indicate that GO enrichment is critically important in COAD patients and is strongly associated with metabolism, especially in the G1/S transition of the mitotic cell cycle.

The relevance between the gene expression in the risk score DFS model and six types of infiltrating immunocytes (B-cells, neutrophils, macrophages, and dendritic cells, CD4+ and CD8+ T-cells) was examined further. The DFS analysis indicated that 3/4 of the expression levels in the risk score model were obviously related with the infiltrating of B-cells, CD8+ T-cells, and macrophages (**Supplementary Figure 2**). The EDAR coefficient in the risk score of DFS was positive and EDAR expression was positively associated with the levels of B-cell and CD8+ T-cell infiltrating in COAD (**Supplementary Figure 2B**). Our analysis indicates that CD8+ T-cells and B-cells may be obviously associated with COAD.

We also investigated the correlation between the gene expression in the risk score OS model expression and 6 types of infiltrating immunocytes. The OS analysis indicated that the genes in the risk score OS model were associated with CD8+ T and B-cell infiltrating. The coefficients of FUT1 and SPP1 in the risk score of OS were positive, and FUT1 expression was critically correlated with the levels of B cell infiltrating in COAD (**Supplementary Figure 3C**). SPP1 expression was critically associated with the infiltrating degrees of CD8+ T-cells in COAD (**Supplementary Figure 3F**). Above all, we considered that CD8+ T-cells and B-cells may be obviously correlated with COAD in both DFS and OS.

Independent Prognostic Role of the Prognostic Model

Kaplan-Meier (KM) survival analysis was conducted for the risk score. The optimal risk score cut-off point in 222 DFS patients with DFS status and time records was 0.18 as illustrated in **Figure 2A**. Based on this best cut-off point, we divided the 222 DFS patients into high- and low-risk groups, the distribution of which was illustrated in **Figure 2A**. The survival curves of the 222 DFS patients under the optimal cut-off point were depicted in **Figure 2B** ($P < 0.0001$). The optimal cut-off point for the risk score in all 214 patients with stages I, II, and III was 0.13, as illustrated in **Figure 2C**. Based on this optimal cut-off point, we separated all 214 DFS patients into high- and low-risk groups, the distribution of which was depicted in **Figure 2C**. The survival curves of the 214 patients are shown in **Figure 2D** ($P < 0.0001$). The survival curves of the patients with stage I, I + II, and III are depicted in **Figures 2E–G** with $P = 0.015$, $P < 0.0001$, and $P = 0.00038$, respectively. All survival curves indicated that the risk score was an independent prognostic parameter, and the low-risk group had a favorable prognosis. As an independent prognostic parameter, the predictive performance of the risk score was represented by the ROC in **Figure 2H**. The AUCs of 1-, 3-, and 5-year predicted by the risk score were 0.779, 0.798, and 0.845, respectively. The OS results for TCGA are included in **Supplementary Figure 4**. The survival curves of DFS and OS for the validation data set are also included in the **Supplementary Figures 5**, **6**.

The Prognostic Model as an Independent Clinical Parameter

Univariate Cox regression analysis of the 214 DFS patients included in the TCGA-COAD data set indicated that sex and our prognostic model were significantly correlated. Multivariate Cox regression analysis revealed that this prognostic model was an independent prognostic parameter of DFS (**Figure 3A**). Among the 554 OS patients included in the TCGA-COAD data set, the prognostic model was also shown to be an independent prognostic parameter of OS (**Figure 3C**), which was consistent with the results from the TCGA-COAD DFS cohort. To make the proposed approach more practical for clinicians, nomograms for both DFS and OS were developed to predict survival probability (**Figures 3B,D**). The C-indices of DFS and OS were 0.786 and 0.748, separately. The results for the validation data set are included in the **Supplementary Figure 7**.

Validation of the Nomogram

The calibration curve for TCGA-COAD was plotted to validate the predictive performance of the nomogram, which was presented in **Figure 4**. The calibration curves illustrated in **Figures 4A–C** represent the 1-, 3-, and 5-year survival of DFS patients in TCGA. The x-axis suggests the nomogram prediction probabilities and the y-axis indicates the observed

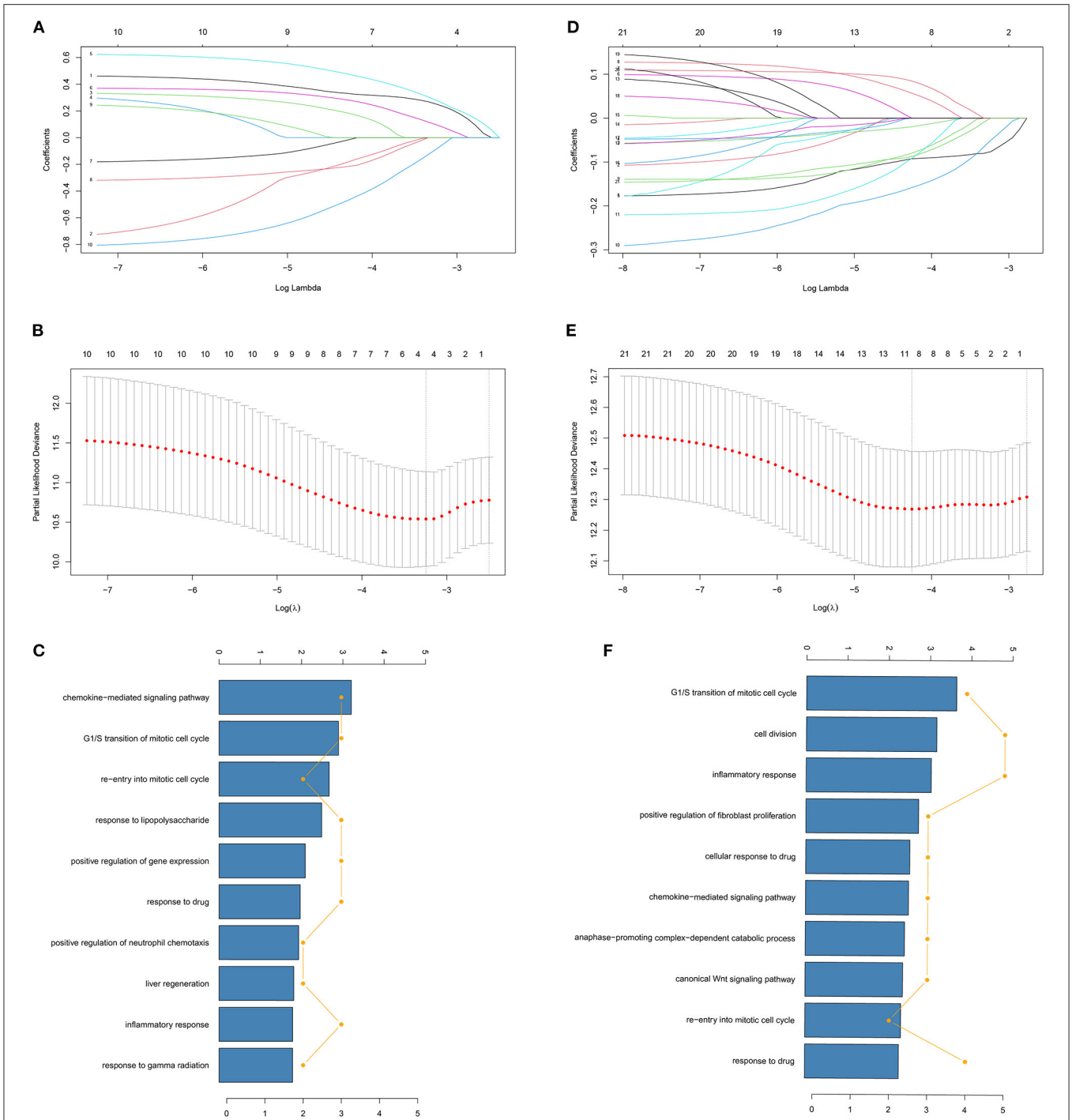


FIGURE 1 | Identification of genes screened via LASSO cox regression analysis. **(A,D)** Coefficients of different genes screened via LASSO cox regression analysis with different λ in DFS and OS. The two dashed lines represent lambda.min and lambda.1se and lambda.min was selected for both DFS and OS. **(B,E)** Partial likelihood deviance of cross-validation with different λ in DFS and OS. **(C,F)** GO analysis of identified genes of LASSO cox regression analysis in DFS and OS.

rates. The dashed line denotes the reference line. The closer the calibration curve was to the reference line, the better the predictive performance of the nomogram. The same results for OS patients in TCGA were illustrated in **Figures 4D,E**. The calibration curves for the validation dataset of GSE39582 were shown in **Supplementary Figure 8**.

In DCA, the y-axis denotes the net benefits and the x-axis indicates the threshold probability. The gray diagonal line in

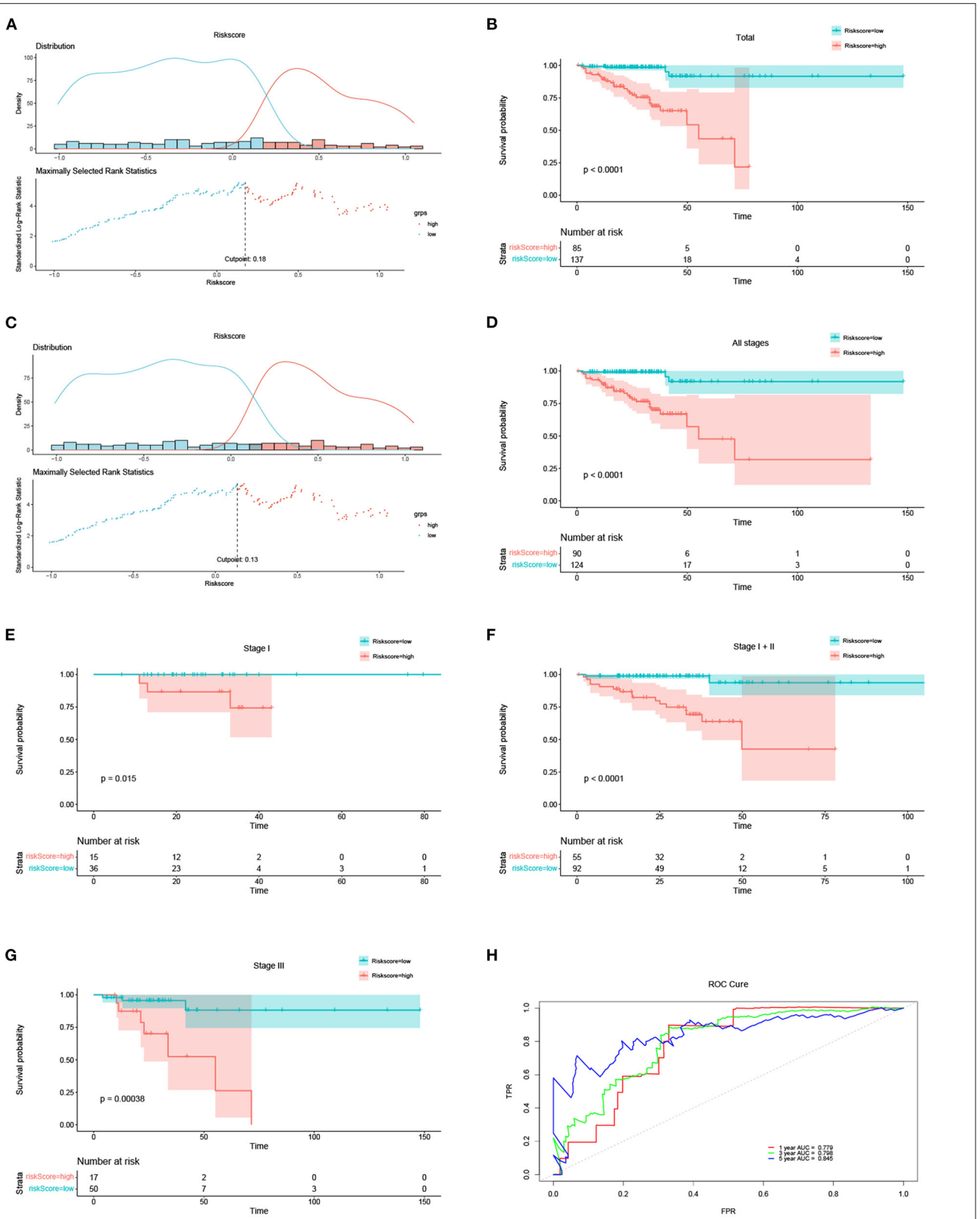


FIGURE 2 | The risk score results acted as an independent prognostic factor in DFS. **(A)** Optimal cut-off of the total 222 DFS patients and distribution of high- and low-risk groups based on the optimal cut-off of 0.18. **(B)** Survival curves of the total 222 DFS patients under the optimal cut-off of 0.18. **(C)** Optimal cut-off of the total 214 DFS patients at different stages (including stage I, II and III) and distribution of high- and low-risk groups based on the optimal cut-off of 0.13. **(D)** Survival curves of all 214 DFS patients under the optimal cut-off of 0.13. **(E–G)** Survival curves of the DFS patients at stages I, I + II, and III, respectively. **(H)** ROC of 1-, 3-, and 5-years predicted by the risk score.

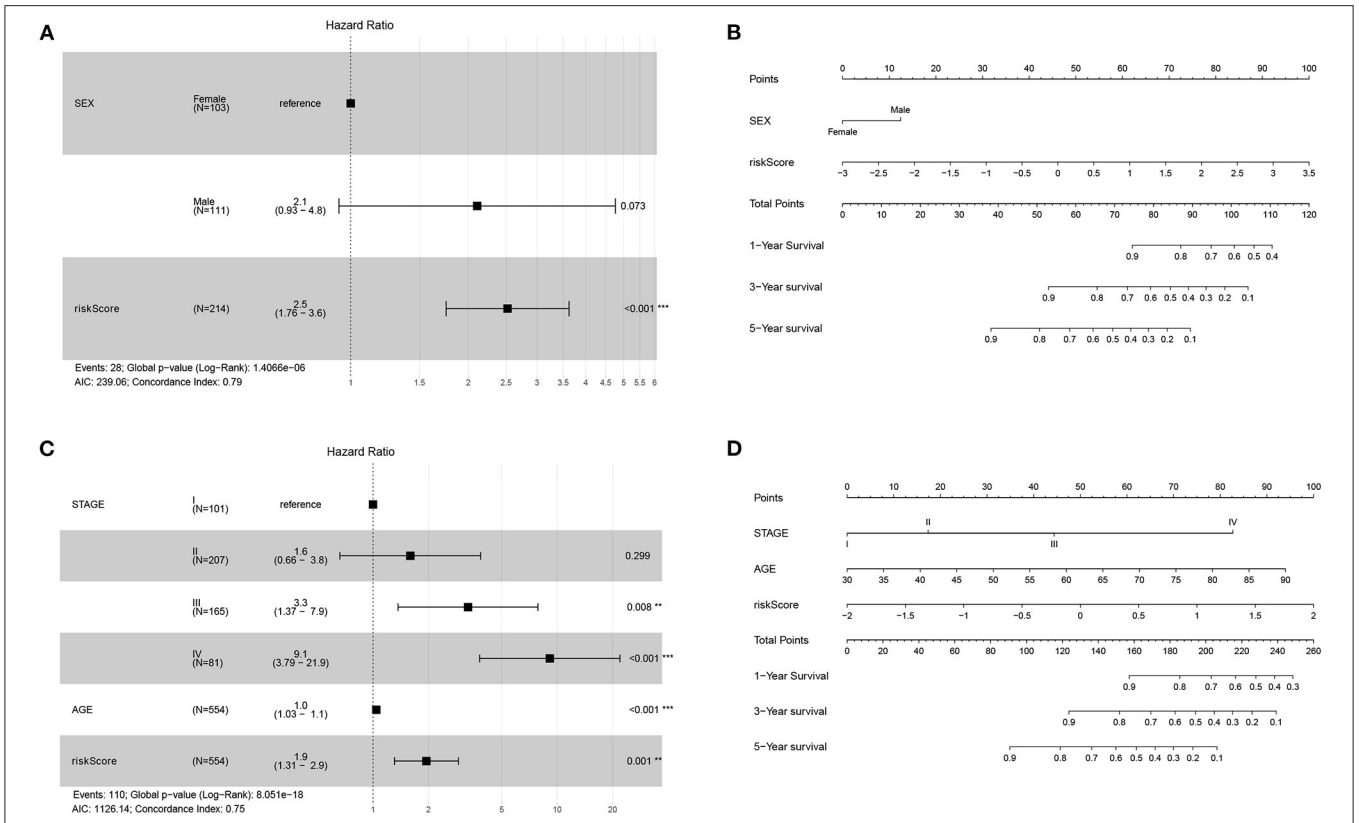


FIGURE 3 | Forest plots and nomograms for DFS and OS in TCGA-COAD. **(A)** Forest plot for DFS in TCGA-COAD. **(B)** Nomogram for DFS in TCGA-COAD. **(C)** Forest plot for OS in TCGA-COAD. **(D)** Nomogram for OS in TCGA-COAD. $^{**}P < 0.01$; $^{***}P < 0.001$.

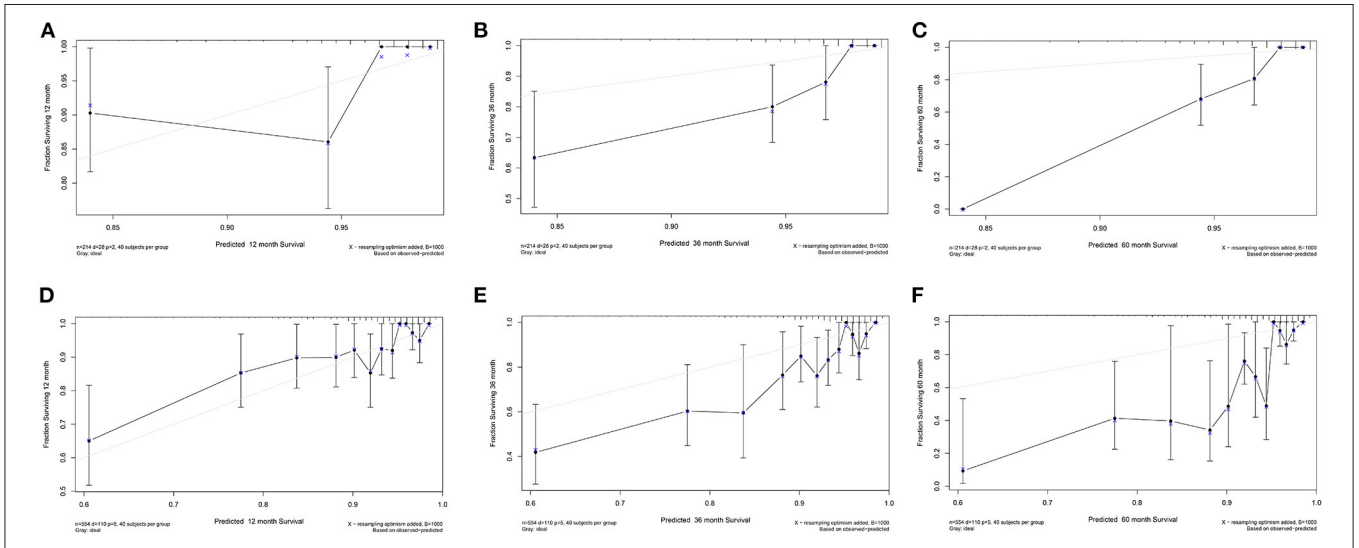


FIGURE 4 | Calibration curve for TCGA-COAD. **(A–C)** Calibration curve of 1-, 3-, and 5-year survival for DFS patients in TCGA. **(D–F)** Calibration curve of 1-, 3-, and 5-year survival for OS patients in TCGA.

Figure 5 denotes the hypothesis that all patients have 1-, 3-, and 5-year survival rates. The black horizontal solid lines represent the surmise that no patients have 1-, 3-, or 5-year survival rate. The 1-, 3-, and 5-year DCA for DFS patients in TCGA were illustrated in **Figures 5A–C**. The higher the net benefit, the better the nomogram. The same results for OS patients in TCGA were

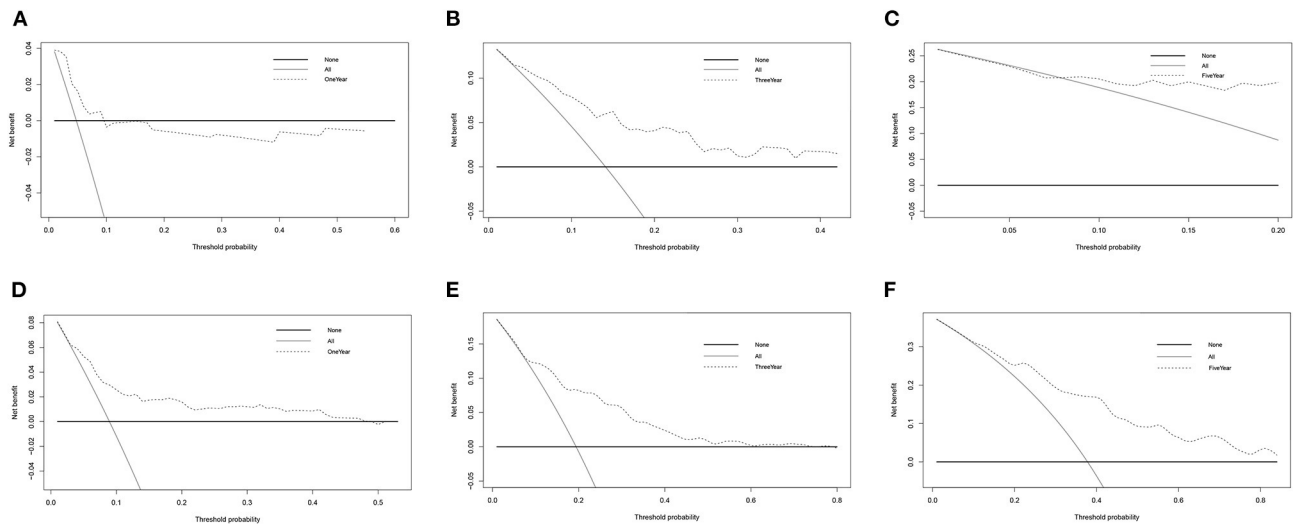


FIGURE 5 | DCA for TCGA-COAD. **(A–C)** DCA of 1-, 3-, and 5-year survival for DFS patients in TCGA. **(D–F)** DCA of 1-, 3- and 5-year survival for OS patients in TCGA.

shown in **Figures 5D–F**. Additionally, the DCA for the validation dataset of GSE39582 is shown in **Supplementary Figure 9**.

External Validation Using the Online Database

In the cBioportal for Cancer Genomics website, FUT1 exhibited the most frequent genetic alterations (2.9%) among the four genes of the prognostic model, with missense mutation being the most common alteration (**Figure 6A**). Consistent with our results, CCND1, FUT1, and PPAT were found to be significantly overexpressed in tumors, whereas EDAR was not significantly expressed in COAD in Oncomine (**Figure 6B**). FUT1 was also found to be significantly overexpressed in COAD tumors compared to normal tissues in the GEPIA database (23) (**Figure 6C**). Survival analyses were carried out by the GEPIA database. Patients with high FUT1 expression displayed remarkably shorter DFS and OS in GEPIA (**Figures 6D,E**); however, CCND1, EDAR, and PPAT showed no significant differences in DFS and OS (**Supplementary Figure 10**). Further, FUT1 was found to be consistent with the DFS and OS results of GSE17536 in PrognoScan (**Figures 6F,G**). Taken together, our findings indicated that FUT1 was an aberrantly expressed gene and high FUT1 expression could be used as a predictor of adverse outcomes.

si-FUT1 Inhibited Cell Migration and Induced Apoptosis in Colon Cancer Cells

A flow cytometry assay was performed to examine the effect of FUT1 on cell apoptosis and the cell cycle. Transfection with si-FUT1 significantly raised the apoptotic rate of HCT-116 cells compared with that of NC cells (**Figures 7A,B**), and significantly reduced the number of cells in the G2 phase (**Figures 7C,D**). Next, cell migration was detected using the streak method. Transfection with si-FUT1 significantly inhibited the cell migration rate of HCT-116 cells compared to that of NC cells (**Figures 7E,F**).

DISCUSSION

Metabolic reprogramming is considered as a new critical feature of malignancy (1). Some potential molecules regulating abnormal metabolism have been tested in preclinical or clinical investigations (24). The interactions between bile acids, cholesterol metabolites, and colonic epithelial cells may be relevant in colon carcinogenesis (25). Meanwhile, big data analysis is an important tool for Prediction task (26–30), and disease forecasting is a vital part of the medical research (31, 32). However, the role and mechanisms of metabolic-related genes in COAD remain unclear. The present study was thus conducted to further explore the association of metabolic genes with prognosis in patients with COAD.

Big data and machine learning algorithms have been widely used in basic and clinical medicine research, such as antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset, and deep neural networks to predict diabetic retinopathy (32, 33). High-throughput multimodal massive medical data is just right for research with big data and machine learning algorithms. The blockchain-enabled internet of medical things (IoMT) can provide strong trust establishment and ensure the traceability of data sharing in the IoMT networks (34). Due to the emergence of heterogeneous IoMT, large volumes of patient data are dispatched to central cloud servers for disease analysis and diagnosis (29). Many publicly available data resources and analytical tools are available for research. Analysis methods including survival analysis, nomogram and data visualization have also been developed with corresponding R packages for researchers to use. Our work used LASSO machine learning algorithms as survival analysis and

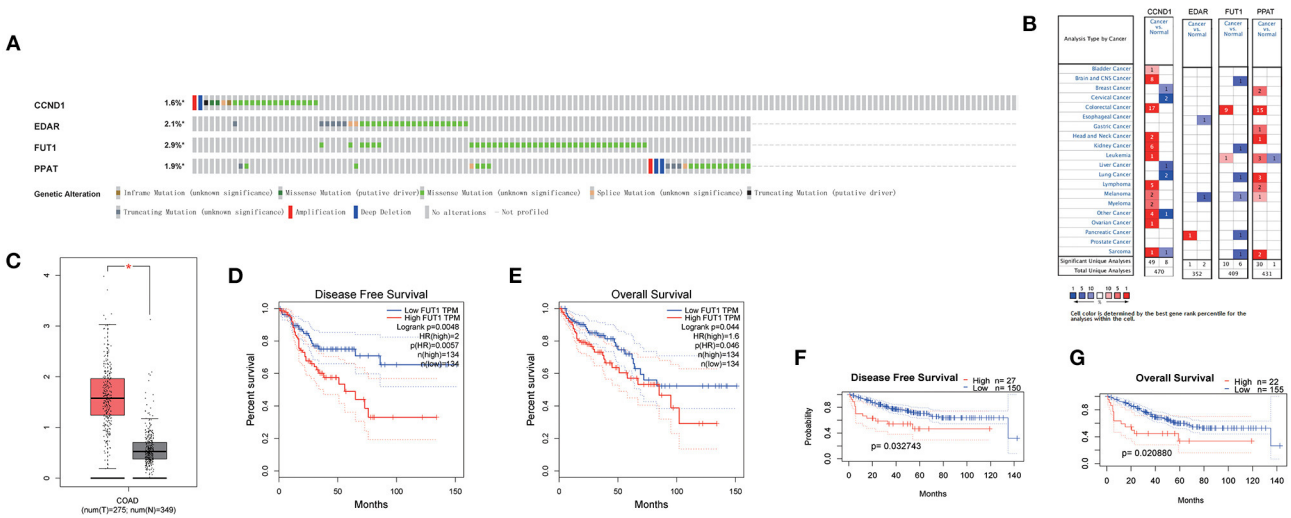


FIGURE 6 | Expression and genetic alterations of the four predictive genes. **(A)** Genetic alterations of the four genes. The data were obtained from the cBioportal in COAD. **(B)** Expression profiles of the four genes in the Oncomine database. **(C)** FUT1 expression in COAD and normal tissue in GEPIA. **(D)** Prognostic value of FUT1 expression in DFS in GEPIA. **(E)** Prognostic value of FUT1 expression in OS in GEPIA. **(F)** Prognostic value of FUT1 expression in DFS in GSE17536 in PrognoScan. **(G)** Prognostic value of FUT1 expression in OS in GSE17536 in PrognoScan. *$P < 0.05$.

prediction tool. Multiple data sources and analytical tools in a way that leverages data from previous studies and yields valuable results even when clinical trials are not yet readily available. As more data resources and analysis tools are developed, we believe that big data and machine learning algorithms will be more widely used in medical research.

In our research, we comprehensively probed the expression of 62 metabolic genes in COAD tumor tissues and the relationships between recurrent patients and DFS and OS. A new prognostic model including four metabolic genes associated with DFS was first established based on the TCGA dataset and validated using the GSE39582 dataset. The validity of the novel signature was shown in the training, validating, and stage subgroups. The signature exhibited a robust prognostic capacity, especially for the short-term survival of patients with COAD. Besides, the OS of patients with high-risk was shorter than that of patients with low-risk in the TNM stage. Multivariate Cox regression analysis indicated that our prognostic model was an independent prognostic parameter for DFS and OS (HR $>$ 1, $P < 0.001$). ROC curve analysis verified the predictive power of the features. Additionally, the forest plot suggested that the risk score was an independent parameter in the multivariate Cox regression analyses. Moreover, a nomogram was built to predict 1-, 3-, and 5-year DFS rates. The efficacy of the nomogram was verified in validation cohorts, and the calibration plots and DCA showed that the precision of the nomogram was good. Therefore, our nomogram may provide simple and accurate prognostic predictions for COAD.

In the cBioportal for Cancer Genomics website, FUT1 possessed the most frequent genetic alterations (2.9%) among the four genes of the prognostic model. FUT1 was also found to be obviously overexpressed at the mRNA level in COAD tumors compared to normal tissues in the GEPIA database (**Figure 6C**). Patients with high FUT1 expression displayed remarkably shorter DFS and OS in the GEPIA to low FUT1 expression (**Figures 6D,E**). Moreover, the FUT1 results were consistent with the DFS and OS results in GSE17536 in PrognoScan (**Figures 6F,G**). Taken together, our findings indicated that FUT1 was an aberrantly expressed gene and therefore high FUT1 expression could be used as a predictor of adverse outcomes.

FUT1 encodes a Golgi stack membrane protein that participates in the synthesis of a precursor of the H antigen. Gene expression profiling analysis unexpectedly showed a significant number of up- and down-regulated metabolism-associated FUT1 genes in nasopharyngeal carcinoma (35). Additionally, FUT1 plays an important role in a regulatory mechanism involving fucosylation through which glucose restriction promotes cancer stemness to drive tumor recurrence and drug resistance. FUT1 overexpression is a poor prognostic indicator of hepatocellular carcinoma (36). We also examined the function of FUT1 in cell apoptosis and cell cycle and performed a flow cytometry assay. Transfection with si-FUT1 significantly raised the apoptotic rate of HCT-116 cells and obviously reduced the number of cells in the G2 phase while also significantly inhibiting the cell migration rate of HCT-116 cells compared to that of NC cells. Therefore, si-FUT1 inhibited cell migration and induced apoptosis in colon cancer cells. FUT1 overexpression may thus be a new prognostic marker and therapeutic target of COAD tumors.

Functional analyses also revealed enriched metabolism-related pathways. The top categories were critically important in COAD patients and strongly associated with metabolism, particularly the G1/S transition of the mitotic cell cycle. Correlation analysis between the genes in the risk score DFS model expression and six types of infiltrating immunocytes suggested that CD8+ T-cells and B-cells may be significantly related with COAD in both DFS and OS.

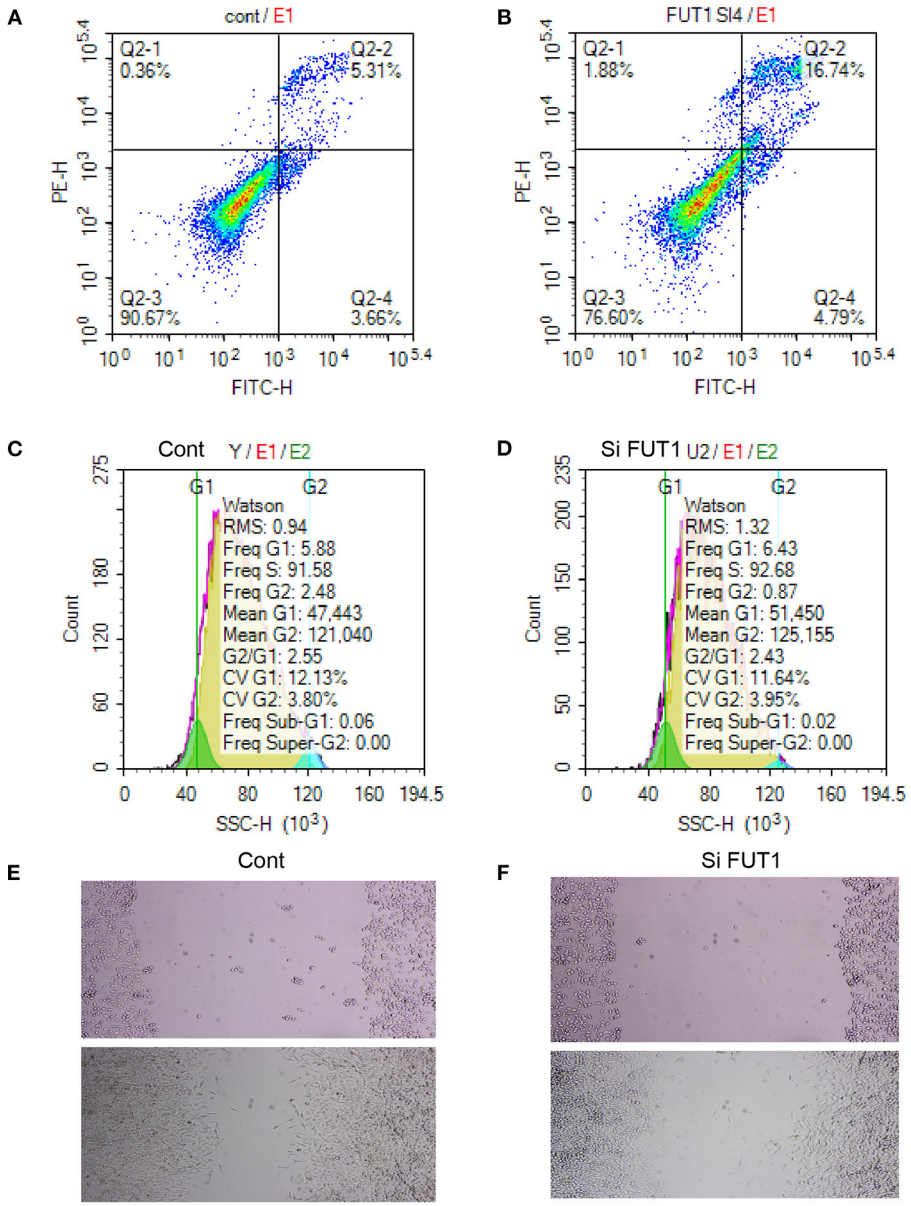


FIGURE 7 | Apoptosis and cycle assay and cell transfection experiment. **(A)** Flow cytometry assay in NC cells. **(B)** Flow cytometry assay in HCT-116 cells. **(C)** Cell cycle in NC cells. **(D)** Cell cycle in HCT-116 cells. **(E)** Cell migration in NC cells. **(F)** Cell migration in HCT-116 cells.

Immunocytes are essential ingredients of the tumor microenvironment. These immunocytes differentiate into subsets with distinct effects, and metabolic reprogramming participates in this process (37). These immunocytes in the tumor microenvironment have metabolic characteristics that differ from those in non-tumor tissues (38). OS was dramatically worse in patients with low levels of CD8+ T-cell infiltrating than those with high levels of CD8+ T-cell infiltrating. The survival rate in patients with high CD8+ T cell infiltrating was 100%. Peritumoral CD8+ T-cell infiltration has an anti-tumor effect in patients with colorectal cancer (39). CD8+ T cell expansion and function rely on glycolysis; however, the mechanisms underlying CD8+ T cell metabolism remain unclear (40). We previously demonstrated that increasing B-cell a infiltrating, clonal expansion, and mutational frequency from the cecum to the sigmoid colon were linked to an increasing number of reactive bacterial species (41). Numerous B-cell clones distribute into two broad networks: one includes the blood, bone marrow, spleen, and lung, whereas the other is distributed to digestive tract, including colorectal tissues (42). B-cell clonal lineages is a basis for investigations on tissue-based immunity, including infection, vaccine response, autoimmunity, and tumor (42). Metabolic reprogramming of tumor cells and the tumor microenvironment are up-and-coming as essential characteristics affecting tumor development, metastasis, and response to treatments (43). A better understanding of

metabolic communications among tumor cells, intestinal flora, and immunocyte populations will open new ways for identifying strategies to boost anti-tumoral immune responses in COAD patients.

Although our work has yielded many meaningful results through bioinformatics methods, these results still need to be further validated by preclinical studies and clinical trials. The use of numerous data sources and analytical tools also makes it more difficult to understand the research process and results. The interpretation of the results of big data studies is also an area prone to controversy. Therefore, there is still work to be done in predicting the prognosis of patients with COAD. For example, more patients and clinical characteristics of patients should be included in further studies.

CONCLUSIONS

In summary, we identified a novel signature comprised of four metabolic genes that could precisely predict the prognosis of patients with COAD. Metabolic-related signatures may have a potential role in the anti-tumor process and serve as therapeutic targets for COAD.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CY and XW designed the article. YZ, RW, and CY collected and evaluated the data and wrote the first draft of the manuscript. All authors reviewed the manuscript, contributed to the interpretation of the results, wrote the final draft of the manuscript, and read and approved the final version of the manuscript.

FUNDING

This work was supported by National Key R&D Program of China (2017YFC1308800).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.860381/full#supplementary-material

Supplementary Figure 1 | Study flowchart.

Supplementary Figure 2 | Correlation between the expression of the genes in the risk score model of DFS and the levels of immune cell infiltration in COAD samples, as investigated by TIMER analysis. **(A)** CCND1. **(B)** EDAR. **(C)** FUT1. **(D)** PPAT.

Supplementary Figure 3 | Correlation between the expression of the genes in the risk score model of OS and immune infiltrating levels in COAD samples, as assessed by TIMER analysis. **(A)** CCNB1. **(B)** CDC6. **(C)** FUT1. **(D)** GPD1L. **(E)** MAD2L1. **(F)** MMP1. **(G)** SPP1. **(H)** TKT.

Supplementary Figure 4 | The risk score results acted as an independent prognostic factor in OS of TCGA. **(A)** Optimal cut-off of the total 566 overall survival patients and distribution of high-risk and low-risk groups based on the optimal cut-off of 0.38. **(B)** Survival curves of the 554 overall survival patients under the optimal cut-off of 0.38. **(C–E)** Survival curves of the patients at stage I + II, stage III, and stage IV, respectively.

Supplementary Figure 5 | Survival curves of DFS in the validation data set of GSE39582. **(A)** Survival curves of 532 DFS patients at different stages (stage I, II, III, and IV). **(B)** Survival curves of DFS patients at stage I. **(C)** Survival curves of DFS patients at stage I + II. **(D)** Survival curves of DFS patients at stage III.

Supplementary Figure 6 | OS curves from the GSE39582 validation data set. **(A)** Survival curves of 569 overall surviving patients at different stages (stage I, II, III, and IV). **(B)** Survival curves of disease-free surviving patients at stage III. **(C)** Survival curves of disease-free surviving patients at stage IV.

Supplementary Figure 7 | Forest plots and nomograms for DFS and OS in the GSE39582 validation data set. **(A)** Forest plot for DFS in GSE39582. **(B)** Nomogram for DFS in GSE39582. **(C)** Forest plot for overall survival in GSE39582. **(D)** Nomogram for overall survival in GSE39582.

Supplementary Figure 8 | Calibration curve for the GSE39582 validation data set. **(A–C)** Calibration curve of 1-, 3-, and 5-year survival for disease-free surviving patients in GSE39582. **(D–F)** Calibration curve of 1-, 3-, and 5-year survival for overall surviving patients in GSE39582.

Supplementary Figure 9 | DCA for validation of data set GSE39582. **(A–C)** DCA of 1-, 3- and 5-year survival for disease-free surviving patients in GSE39582. **(D–F)** DCA of 1-, 3-, and 5-year survival for overall surviving patients in GSE39582.

Supplementary Figure 10 | Prognostic value of gene expression in GEPIA. **(A)** CCND1 in DFS. **(B)** CCND1 in OS. **(C)** EDAR in DFS. **(D)** PPAT in OS. **(E)** PPAT in DFS. **(F)** PPAT in OS.

Supplementary Table 1 | List of metabolic-related genes.

REFERENCES

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
2. La Vecchia S, Sebastian C. Metabolic pathways regulating colorectal cancer initiation and progression. *Semin Cell Dev Biol.* (2020) 98:63–70. doi: 10.1016/j.semcdb.2019.05.018
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* (2018) 68:7–30. doi: 10.3322/caac.21442
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
5. Chen J. Trends of cancer incidence and mortality in China. *J Zunyi Med Univ.* (2018) 41:653–62. doi: 10.1093/annonc/mds069
6. Zheng RS, Sun KX, Zhang SW, Zeng HM, Zou XN, Chen R, et al. [Report of cancer epidemiology in China, 2015]. *Zhonghua Zhong Liu Za Zhi.* (2019) 41:19–28. doi: 10.3760/cma.j.issn.0253-3766.2019.01.005
7. Hu M, Fu X, Si Z, Li C, Sun J, Du X, et al. Identification of differently expressed genes associated with prognosis and growth in colon adenocarcinoma based on integrated bioinformatics analysis. *Front Genet.* (2019) 10:1245. doi: 10.3389/fgene.2019.01245
8. Wu Z, Zuo M, Zeng L, Cui K, Liu B, Yan C, et al. Oma1 reprograms metabolism under hypoxia to promote colorectal cancer development. *EMBO Rep.* (2021) 22:e50827. doi: 10.15252/embr.202050827

9. Wang Y, Lu JH, Wu QN, Jin Y, Wang DS, Chen YX, et al. Lncrna Linris stabilizes Igf2bp2 and promotes the aerobic glycolysis in colorectal cancer. *Mol Cancer.* (2019) 18:174. doi: 10.1186/s12943-019-1105-0
10. Li Z, Chan C. Integrating gene expression and metabolic profiles. *J Biol Chem.* (2004) 279:27124–37. doi: 10.1074/jbc.M403494200
11. Yin C, Zhang J, Shen M, Gu Z, Li Y, Xue W, et al. Matrix metallopeptidase 14: a candidate prognostic biomarker for diffuse large B-cell lymphoma. *Front Oncol.* (2020) 10:1520. doi: 10.3389/fonc.2020.01520
12. Yin C, Zhang J, Guan W, Dou L, Liu Y, Shen M, et al. High expression of Clec11a predicts favorable prognosis in acute myeloid leukemia. *Front Oncol.* (2021) 11:608932. doi: 10.3389/fonc.2021.608932
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
14. Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, et al. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med.* (2018) 22:5928–38. doi: 10.1111/jcmm.13863
15. Li W, Xu L, Han J, Yuan K, Wu H. Identification and validation of tumor stromal immunotype in patients with hepatocellular carcinoma. *Front Oncol.* (2019) 9:664. doi: 10.3389/fonc.2019.00664
16. The Gene Ontology C. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res.* (2019) 47:D330–8. doi: 10.1093/nar/gky1055
17. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. Timer: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* (2017) 77:e108–10. doi: 10.1158/0008-5472.CAN-17-0307
18. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. Oncomine: a cancer microarray database and integrated data-mining platform. *Neoplasia.* (2004) 6:1–6. doi: 10.1016/S1476-5586(04)80047-2
19. Wilson BJ, Giguere V. Identification of novel pathway partners of P68 and P72 Rna helicases through oncomine meta-analysis. *BMC Genomics.* (2007) 8:419. doi: 10.1186/1471-2164-8-419
20. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The Cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* (2012) 2:401–4. doi: 10.1158/2159-8290.CD-12-0095
21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Sci Signal.* (2013) 6:pl1. doi: 10.1126/scisignal.2004088
22. Mizuno H, Kitada K, Nakai K, Sarai A. Prognoscan: a new database for meta-analysis of the prognostic value of genes. *BMC Med Genomics.* (2009) 2:18. doi: 10.1186/1755-8794-2-18
23. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. Gepia: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* (2017) 45:W98–102. doi: 10.1093/nar/gkx247
24. Piccinin E, Villani G, Moschetta A. Metabolic aspects in Nafld, nash and hepatocellular carcinoma: the role of Pgc1 coactivators. *Nat Rev Gastroenterol Hepatol.* (2019) 16:160–74. doi: 10.1038/s41575-018-0089-3
25. Reddy BS, Wynder EL. Metabolic epidemiology of colon cancer. Fecal bile acids and neutral sterols in colon cancer patients and patients with adenomatous polyps. *Cancer.* (1977) 39:2533–9.
26. Naeem A, Javed AR, Rizwan M, Abbas S, Lin JCW, Gadekallu TR. Dare-Sep: a hybrid approach of distance aware residual energy-efficient Sep for Wsn. *IEEE T Green Commun.* (2021) 5:611–21. doi: 10.1109/TGCN.2021.3067885
27. Gadekallu TR, Alazab M, Kaluri R, Maddikunta PKR, Bhattacharya S, Lakshmanna K, et al. Hand gesture classification using a novel Cnn-Crow search algorithm. *Complex Intell Syst.* (2021) 7:1855–68. doi: 10.1007/s40747-021-00324-x
28. Kumar R, Tripathi R, Marchang N, Srivastava G, Gadekallu TR, Xiong NN. A secured distributed detection system based on Ipfs and blockchain for industrial image and video data security. *J Parallel Distr Com.* (2021) 152:128–43. doi: 10.1016/j.jpdc.2021.02.022
29. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and Puf-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet Things J.* (2021) 8:1. doi: 10.1109/JIOT.2021.3117762
30. Wang W, Fida MH, Lian Z, Yin Z, Pham QV, Gadekallu TR, et al. Secure-enhanced federated learning for AI-empowered electric vehicle energy prediction. *IEEE Consumer Electr Mag.* (2021) 10:1. doi: 10.1109/MCE.2021.3116917
31. Agrawal S, Sarkar S, Srivastava G, Maddikunta PKR, Gadekallu TR. Genetically optimized prediction of remaining useful life. *Sustain Comput Infor.* (2021) 31:100565. doi: 10.1016/j.suscom.2021.100565
32. Chowdhary CL, Reddy GT, Parameshachari BD. *Computer Vision and Recognition Systems : Research Innovations and Trends*. First ed. Palm Bay, FL: Apple Academic Press, Inc. (2022).
33. Reddy GT, Bhattacharya S, Ramakrishnan SS, Chowdhary CL, Hakak S, Kaluri R, et al., editors. An ensemble based machine learning model for diabetic retinopathy classification. In: *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*. Vellore (2020).
34. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based Ecdsa with fault-tolerant batch verication protocol for blockchain-enabled Iomt. *IEEE J Biomed Health Inform.* (2021) 25:1. doi: 10.1109/JBHI.2021.3112693
35. Ye Y, Zhou Y, Zhang L, Chen Y, Lyu X, Cai L, et al. Ebv-Mir-Bart1 is involved in regulating metabolism-associated genes in nasopharyngeal carcinoma. *Biochem Biophys Res Commun.* (2013) 436:19–24. doi: 10.1016/j.bbrc.2013.05.008
36. Loong JH, Wong TL, Tong M, Sharma R, Zhou L, Ng KY, et al. Glucose deprivation-induced aberrant fut1-mediated fucosylation drives cancer stemness in hepatocellular carcinoma. *J Clin Invest.* (2021) 131:e143377. doi: 10.1172/JCI143377
37. Biswas SK. Metabolic reprogramming of immune cells in cancer progression. *Immunity.* (2015) 43:435–49. doi: 10.1016/j.immuni.2015.09.001
38. Xiao Z, Dai Z, Locasale JW. Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat Commun.* (2019) 10:3763. doi: 10.1038/s41467-019-11738-0
39. Funada Y, Noguchi T, Kikuchi R, Takeno S, Uchida Y, Gabbert HE. Prognostic significance of Cd8+ T cell and macrophage peritumoral infiltration in colorectal cancer. *Oncol Rep.* (2003) 10:309–13. doi: 10.3892/or.10.2.309
40. Hu Z, Qu G, Yu X, Jiang H, Teng XL, Ding L, et al. Acylglycerol kinase maintains metabolic state and immune responses of Cd8(+) T cells. *Cell Metab.* (2019) 30:290–302 e5. doi: 10.1016/j.cmet.2019.05.016
41. James KR, Gomes T, Elmentaite R, Kumar N, Gulliver EL, King HW, et al. Distinct microbial and immune niches of the human colon. *Nat Immunol.* (2020) 21:343–53. doi: 10.1038/s41590-020-0602-z
42. Meng W, Zhang B, Schwartz GW, Rosenfeld AM, Ren D, Thome JJC, et al. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol.* (2017) 35:879–84. doi: 10.1038/nbt.3942
43. Ngwa VM, Edwards DN, Philip M, Chen J. Microenvironmental metabolism regulates antitumor immunity. *Cancer Res.* (2019) 79:4003–8. doi: 10.1158/0008-5472.CAN-19-0617

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zheng, Wu, Wang and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpubh.2022.847252

Deep Data Analysis-Based Agricultural Products Management for Smart Public Healthcare

Wenjing Yan[1], *Zesheng Zhang*[1], *Qingchuan Zhang*[1], *Ganggang Zhang*[2], *Qiaozhi Hua*[3]* *and Qiao Li*[4]

[1] National Engineering Laboratory for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing, China, [2] Digital Campus Construction Center, Capital Normal University, Beijing, China, [3] Computer School, Hubei University of Arts and Science, Xiangyang, China, [4] Chongqing Key Laboratory of Intelligent Perception and Blockchain Technology, Chongqing Technology and Business University, Chongqing, China

Agricultural is an indispensably public healthcare industry for human beings at any time and smart management of it is of great significance. Since substantial technical advance relies on long-term efforts and continuous progress, reasonably scheduling the distribution of agricultural products acts as a key aspect of smart public healthcare. The most intuitive factor affecting the distribution of agricultural products is its dynamic price. Forecasting price fluctuations in advance can optimize the distribution of agricultural products and pave the way to smart public healthcare. Most researchers study the prices of various agricultural products separately, without considering the interaction of different agricultural products in the time dimension. This study introduces a typical deep learning model named graph neural network (GNN) for this purpose and proposes deep data analysis-based agricultural products management for smart public healthcare (named GNN-APM for short). The highlight of GNN-APM is to take latent correlations among multiple types of agricultural products into consideration when modeling evolving rules of price sequences. A case study is set up with the use of real-world data of the agricultural products market. Simulative results reveal that the designed GNN-APM functions well.

Keywords: graph neural network, agricultural products, public healthcare, deep data analysis, smart management

OPEN ACCESS

Edited by:
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Durai Raj Vincent P. M.,
VIT University, India
Heng Wang,
Henan Agricultural University, China
Shizheng Zhang,
Zhengzhou University of Light Industry, China
Weizheng Wang,
Cloud County Community College, United States

***Correspondence:**
Qiaozhi Hua
11722@hbuas.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *01 January 2022*
Accepted: *21 February 2022*
Published: *07 April 2022*

Citation:
Yan W, Zhang Z, Zhang Q, Zhang G, Hua Q and Li Q (2022) Deep Data Analysis-Based Agricultural Products Management for Smart Public Healthcare. Front. Public Health 10:847252. doi: 10.3389/fpubh.2022.847252

1. INTRODUCTION

Since ancient times, agriculture has been a life industry for human survival, which is closely related to the most basic life guarantee of human beings. At present, food shortage is one of the most important problems faced by many regions in the world (1). This phenomenon is generally reflected in two aspects (2). For one thing, there is still room for improvement in current agricultural technology, which makes grain yield fail to meet the expectations (3). For another, due to the lack of scientific management and scheduling strategy, the production of agricultural products is not reasonable (4). The exploration of advanced agricultural technology has lasted for at least a 100 years, and some technological breakthroughs have been made in some key fields (5, 6). However, this process exerts an imperceptible influence as we all know, which needs to be accumulated over a long period of time to make progress (7). The use of advanced computing technology to manage the agricultural products market can improve the distribution and dispatch efficiency of global agricultural products to a certain extent (8), and then alleviate the problem of food shortage (9). The key to improving management efficiency is to forecast the market conditions of several major types

of agricultural products (10). To realize such a goal, data-driven methods are the most intuitive ways (11, 12).

Rationalizing the distribution of agricultural products is an important aspect of smart public healthcare. Predicting price fluctuations in advance can optimize the layout of agricultural products and pave the way for smart public healthcare. Many scholars have studied the agricultural products market in recent years (13–32). For example, Fan et al. (14) analyzed the value-added mechanism of agricultural products circulation value chain and put forward three optimization methods of agricultural products organization mode. These existing researches are mainly realized through the research methods of social sciences. Their analyses focus on the mechanism of social development and evolution but have to deal with large-scale computational tasks in a manual way. Without the assistance of intelligent computing, these methods are always faced with certain limitations. There are also some scholars who use intelligent computing methods to solve this problem. They regard the forecasting problem of the future agricultural products market as a time series forecasting problem based on historical data. However, most of them just treat different agricultural products as independent categories and then build time series prediction models separately. In this way, potential relationships between categories are ignored, so that the precision of the modeling process is reduced.

In order to solve the above challenges, a graph neural network (GNN) can be used to model this time series prediction problem (33). Different from the traditional model that deals with grid-structured data, GNN deals with the data of topological structure. GNN is based on deep learning (34) and is widely used in various fields due to its good performance and interpretability (35). It can deeply perceive the relationship between entities in the process of modeling by graph-structured data. Vertexes and edges connecting vertexes together constitute graph structure. The vertexes are object entities, and the edges are the specific relationship between the entities. Specifically, several common agricultural products can be regarded as entities, and the correlation between them can be regarded as edges, which together constitute a kind of graph network. By introducing a neural computing structure, a GNN model for time series prediction can be constructed. Therefore, this study designs a graph neural network-based smart management for the agricultural products market (GNN-APM). The main highlights of this article can be summed up as follows:

- The internal complexity of agricultural products systems is investigated for further management.
- The GNN is employed to construct a time-series price prediction method for agricultural products systems.
- A case study is carried out to evaluate the performance of the proposed method on real-world scenes.

2. SYSTEM MODEL

The left part of **Figure 1** illustrates the learning and training process inside GNN-APM. First, the feature space of different types of agricultural products is coded into vectorized representations. Then a prediction model is obtained by learning the data samples. As a necessity in our life, agricultural products have a strong correlation among various categories, and these correlations are very meaningful and worth exploring further. The complex interrelation between different categories affects their demand to a great extent. The introduction of the GNN model deeply excavates the correlation between different categories of agricultural products and builds a graph structure with agricultural products as the entity. Through analysis, the market conditions of agricultural products can be predicted and further intelligent management of public health can be realized.

Generalized to the problem scenario in this study, there are several types of agricultural products whose market conditions need to be forecasted. The learning algorithm of GNN-APM is shown in **Algorithm 1**. Types of agricultural products are viewed as the set of nodes, and their internal relations are regarded as the set of edges. Market conditions of agricultural products refer to the average market price in this research and will be updated temporally. Each time that the market condition is updated, is defined as a timestamp t which ranges from 1 to M. During each timestamp, market conditions of all the agricultural types are denoted as $V_i^{(t)}$, where i is the index number of agricultural products types. Inputting market conditions data of M timestamps, the main goal is to predict unknown market conditions data of following timestamps. Obviously, internal relations among nodes are likely to influence the tendency of market conditions. Thus, a relationship-aware sequential forecasting problem is formulated, and the GNN model is adopted to deal with such a problem.

Algorithm 1 | The learning algorithm of GNN-APM.

Input: The market condition dataset: D; node set: V; number of agricultural products: N; total timestamp: M; learning rate: l; parameter set: Θ; penalty parameter: p;

initial $iter = 1$
repeat
  for $t \in [1, M]$ **do**
    for $i \in [1, N]$ **do**
      compute $V_i^{(t)} = f(\theta(;;id(V_i); \Theta)$;
      compute $loss = \sum_{i=1}^{N} \sum_{t=1}^{M} [\| V_i^{(t)} - \hat{V}_i^{(t)} \| + p * \| \Theta \|_F^2]$;
      compute the gradient of Θ according to $loss$;
      update model parameters Θ according to their gradients and learning rate l;
    end for
  end for
  $iter = iter + 1$;
until convergence

3. METHODOLOGY

Graph convolution network (GCN), a typical GNN model, is utilized here to model correlated sample space. The GCN extends the convolution operation to non-Euclidean data with graph

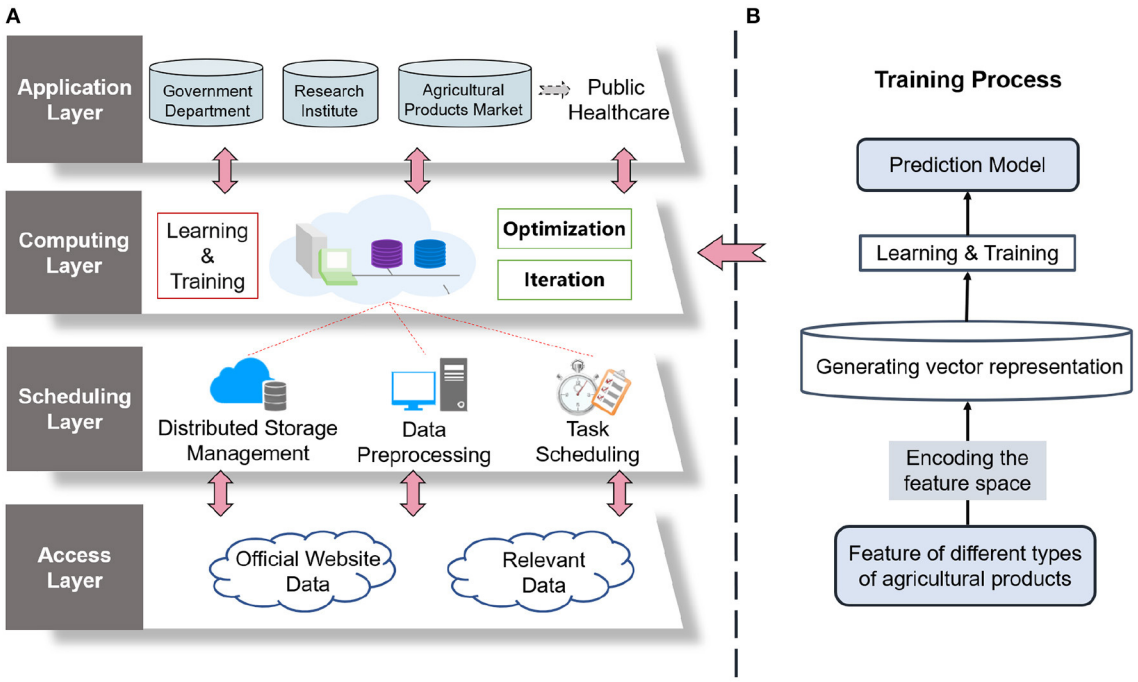


FIGURE 1 | System model and overview of the designed graph neural network-based smart management for agricultural products market (GNN-APM). **(A)** architecture and **(B)** training process in computing layer.

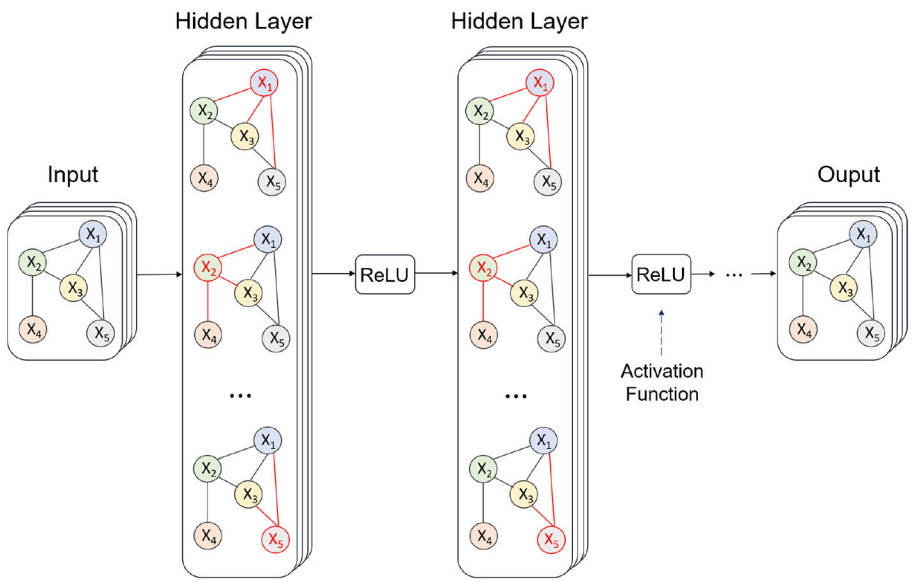


FIGURE 2 | Workflow of the graph convolutional network employed for prediction.

structure. It is a deep learning method for graph structured data. Graph data can naturally represent data structures in real life, such as traffic networks, communication networks, and social networks. In other words, it is the way to represent this kind of data format. Unlike image and text data, graph data has different local structures for each node. This is because the nodes in the graph represent the different entities in the network, and the edges that connect the nodes represent the relationships between the entities.

As is shown in **Figure 2**, taking graph structure as input, GCN obtains new node representation through graph convolution operation on neighbor nodes of each node in the graph. Then,

all nodes are pooled to obtain the representation of the whole graph. In particular, an undirected graph with nodes is defined as $G(V, E)$, where V is the number of nodes and E is the edge between two nodes. Enumerating i from 1 to N, v_i constitutes the node set V. Let j denote the index number of nodes different from node i, edge e_{ij} constitutes all the edges between pairs of nodes. Additionally, all the edge states inside graph G are able to make up an adjacency matrix A.

There are two kinds of GCN methods: spectral method and spatial method. Spectral CNN is the first method to construct a convolutional neural network on the graph. This method uses the convolution theorem on the graph to define graph convolution from the spectral domain. Specifically, it uses the convolution theorem to define the graph convolution operator in each layer. Under the guidance of the loss function, it learns the convolution kernel by gradient backpropagation and builds neural networks by stacking multiple layers. The spectral method is more general in most time-series prediction problems and is selected as the main technique of this study.

The spectral GCN method derives from the Fourier transform (FT) theory which can transform signals of the time domain into signals of the frequency domain. After such transformation, complicated convolution operations of the time domain can be approximated as multiplication operations of the frequency domain. The FT and inverse Fourier transform (IFT) are defined as follows:

$$\hat{x} = FT(x) = U^T x \tag{1}$$

$$x = IFT(x) = U\hat{x} \tag{2}$$

where U is the eigenvector that approximates the FT to matrix computation. Thus, the graph convolution operation is defined as follows:

$$x \otimes_G g = U\left(U^T x \odot U^T g\right) \tag{3}$$

where x is the input, $\otimes_G$ is the graph convolution operator, g is the core, and $\odot$ is the harmand product operator. Introducing Laplacian eigenvector as a basis function, the input signal can be expanded as:

$$x\begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(\gamma) \end{bmatrix} = \left(u_1, u_2, \cdots, u_\gamma\right)\begin{bmatrix} \hat{x}(1) \\ \hat{x}(2) \\ \vdots \\ \hat{x}(\gamma) \end{bmatrix} \tag{4}$$

Expanding g with matrix forms and then substituting Equation (4) into Equation (3), the following formula can be deduced:

$$x \otimes_G g = U\begin{bmatrix} \hat{g}(\lambda_1) & & \\ & \ddots & \\ & & \hat{g}\left(\lambda_\gamma\right) \end{bmatrix} \tag{5}$$

Letting $Chev(\lambda_1)$ denote first-order Chebyshev polynomials of λ_1, the convolution operator g can be approximated as:

$$g_\theta = U\begin{bmatrix} Chev(\lambda_1) & & \\ & \ddots & \\ & & Chev\left(\lambda_\gamma\right) \end{bmatrix} \tag{6}$$

Among, the $Chev\left(\lambda_\gamma\right)$ can be represented as:

$$Chev\left(\lambda_\gamma\right) = \xi_0 \Gamma_0\left(\lambda_\gamma\right) + \xi_1 \Gamma_1\left(\lambda_\gamma\right) \tag{7}$$

Hence, Equation (5) can be rewritten as the following formula:

$$x \otimes_G g = \left[\alpha_0 - \alpha_1\left(E^{-\frac{1}{2}} \cdot A \cdot E^{-\frac{1}{2}}\right)\right]x \tag{8}$$

where E is the degree matrix, and α_0 and α_1 are parameters to be learned. For simplicity, it is supposed to set $\alpha_0 = \alpha_1 = -\theta$. Therefore, the above equation can be rewritten as:

$$x \otimes_G g = \left[\theta\left(E^{-\frac{1}{2}} \cdot A \cdot E^{-\frac{1}{2}} + D_\eta\right)\right] \tag{9}$$

In order to facilitate searching for optimum, renormalization operation is conducted on the above formula:

$$E^{-\frac{1}{2}} \cdot A \cdot E^{-\frac{1}{2}} + D_\eta \approx E^{-\frac{1}{2}} \cdot \tilde{A} \cdot E^{-\frac{1}{2}} \tag{10}$$

$$\tilde{A} \approx A + D_\eta \tag{11}$$

$$\tilde{E}_{ij} \approx \sum_i \tilde{A}_{ij} \tag{12}$$

where $\tilde{E}_{ij}$ is the degree matrix of the i-th node, and $\sum_i \tilde{A}_{ij}$ is the number of edges between the i-th node and other nodes. The final expression of graph convolution operation can be represented as:

$$x \otimes_G g = \theta\left(E^{-\frac{1}{2}} \cdot \tilde{A} \cdot E^{-\frac{1}{2}}\right)x \tag{13}$$

At the t-th timestamp, the main input is related to the outputs of the previous several timestamps and the relation status among nodes. For the i-th node, its information state at the t-th timestamp can be represented as the following formula:

$$S_i^{(t)} = \beta \cdot O_i^{(t)} \cdot W_{S1} + (1 - \beta) \cdot a_i \cdot W_{S2} \tag{14}$$

where $O_i^{(t)}$ is a vector that records output values at previous several timestamps, a_i is a vector that records relation status between the i-th node and other nodes, β is a tuning parameter that adjusts the weight of two parts in Equation (14), and W_{S1} and W_{S2} are parameters to be learned. It is widely known that the internal of GCN is the information propagation process, as GCN emphasizes modeling of various dynamic or static relations. Accordingly, the representative vectors for node status can be also

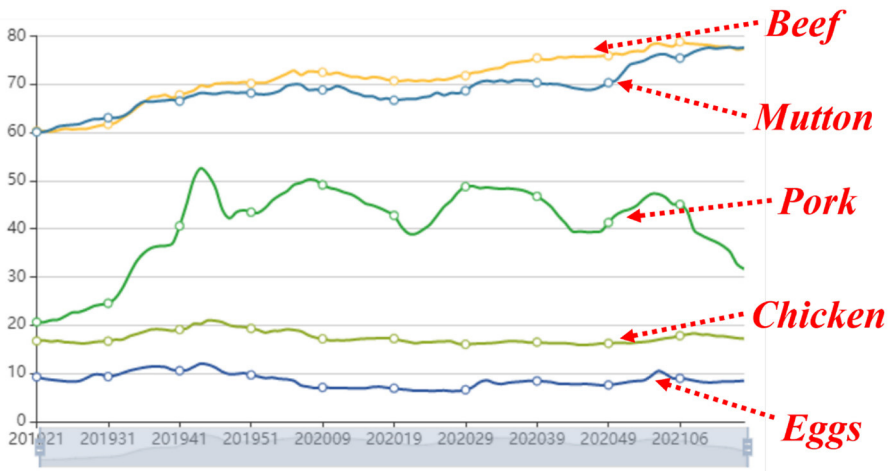


FIGURE 3 | The total tendency of market conditions concerning five types of agricultural products.

propagated to the following timestamps, which can be expressed as the following formula:

$$S_i^{(t+1)} = \sigma_1\left[\tilde{E}^{-\frac{1}{2}} \cdot S_i^{(t)} \cdot \tilde{E}^{-\frac{1}{2}} \cdot W_{S3} + b_{S1}\right] \tag{15}$$

where W_{S3} and b_{S1} are parameters, and $\sigma_1(\cdot)$ is the Reluctant Unit activation function represented as follows:

$$\sigma_1(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases} \tag{16}$$

Hence, the prediction result at the t-th timestamp can be calculated as the following formula:

$$V_i^{(t)} = \sigma_1\left[S_i^{(t)} \cdot W_{S4} + b_{S2}\right] \tag{17}$$

where W_{S4} and b_{S2} are parameters. As for training, the following optimization objective can be formulated to search for the optimal parameters:

$$\min \sum_{i=1}^{N}\sum_{t=1}^{M}\left[\left\|V^{(t)} - \hat{V}^{(t)}\right\| + p \cdot \|\Theta\|_F^2\right] \tag{18}$$

where $V^{(t)}$ is the predicted result at the t-th timestamp, $\hat{V}^{(t)}$ is the real result at the t-th timestamp, Θ is the parameter set, p is the penalty parameter, and $\|\cdot\|_F^2$ is the Frobenius norm. Finally, the Adam optimization algorithm can be utilized to search optimal solution for Equation 18.

4. SETTING OF CASE STUDY

To evaluate the GNN-APM designed in this study, real-world data is used here to set a simulative analysis situation. The real-world data was crawled from the official website of the Ministry of agriculture of China [1], including the market data of several key agricultural products from April 2019 to March 2021. The data demonstrates the average market price of agricultural products wholesale and contains five types of agricultural products. Market condition data for these agricultural products are updated once a week, and there are totally 96 weeks of data concerning the five types of agricultural products. Of all the five types of agricultural products, there are 10 kinds of node combinations, indicating 10 kinds of edges among these nodes. In other words, the adjacency matrix in this situation is a matrix with five columns and five lines, representing relation status between every two combinations of nodes.

During each round of simulations, these 10 groups of relations are randomly generated according to a Gaussian distribution whose mean is set to 0.5 and variance is set to 0.05. As for the ratio between training data and testing data, it is majorly set to 7:3 and 6:4. To quantify the error between prediction results and real results, two typical metrics are selected. They are mean absolute error (MAE) and root mean squared error (RMSE). In addition, two general prediction models are employed as baseline methods for comparison. Different from the designed GNN-APM in this study, the two baseline methods never take internal correlations among nodes into consideration. The two methods are the long short-term memory (LSTM) model and the multi-layer perceptron (MLP) model.

5. RESULTS AND ANALYSIS

In this study, the whole simulative experiments are composed of three parts. First, the fluctuation tendency of the involved five agricultural product types is visualized using a curve diagram. Second, the prediction efficiency of the GNN-APM on five objects is compared with two baseline methods. Third, the

[1]http://www.moa.gov.cn/

TABLE 1 | Mean absolute error (MAE) results when proportion of training data ranges from 50 to 70% and learning rate ranges from 0.001 to 0.002.

Setting	Learning rate: 0.001			Learning rate: 0.002		
	50%	60%	70%	50%	60%	70%
MLP	1.189	1.047	1.016	1.232	1.175	1.143
LSTM	1.116	1.052	1.044	1.139	1.091	1.107
GNN-APM	**0.964**	**0.927**	**0.875**	**0.989**	**0.903**	**1.016**

The bold values correspond to results of the proposed GNN-APM method.

TABLE 2 | Root mean squared error (RMSE) results when the proportion of training data ranges from 50 to 70% and the learning rate ranges from 0.001 to 0.002.

Setting	Learning rate: 0.001			Learning rate: 0.002		
	50%	60%	70%	50%	60%	70%
MLP	1.664	1.458	1.373	1.761	1.663	1.589
LSTM	1.512	1.408	1.387	1.553	1.476	1.481
GNN-APM	**1.289**	**1.245**	**1.177**	**1.343**	**1.184**	**1.353**

The bold values correspond to results of the proposed GNN-APM method.

robustness of the GNN-APM is tested by changing several parameter combinations.

5.1. Data Pre-processing

Figure 3 visualizes the total tendency of market conditions for five types of agricultural products. It can be observed from the figure that five curves from the bottom to the top correspond to eggs, chicken, pork, mutton, and beef. During a long period about nearly 2 years, eggs and chicken remain relatively stable, mutton and beef show an ascending tendency, and the pork fluctuates frequently. These five types of agricultural products possess their own fluctuation tendencies and satisfy the assumption of diversity. And it can be also seen that the fluctuation tendency of pork has some effect on the other four types of agricultural products. Thus, the assumption that correlations exist among these types of agricultural products is reasonable.

5.2. Performance Assessment

Tables 1, **2**, respectively give MAE results and RMSE results of these experimental methods when the proportion of training data ranges from 50 to 70% and the learning rate ranges from 0.001 to 0.002. Each of them has five lines and seven rows. The first two lines list the experimental setting, and the other lines present the experimental results of three methods. The first row lists three experimental methods, the second to the fourth rows present results under a learning rate of 0.001, the fifth to the seventh rows present results under a learning rate of 0.002. It can be observed from the two tables that MAE results and RMSE results of GNN-APM are below two other baseline methods, regardless of the proportion of training data and learning rate. This demonstrates that the performance of the GNN-APM is better than baseline methods.

Figures 4, **5** illustrates prediction efficiency with respect to using two metrics: MAE and RMSE. As there are totally five types of agricultural products involved, the MAE results and RMSE results are obtained as the mean value of prediction results on the five types. This figure has two subfigures, corresponding to MAE results and RMSE results. Among them, **Figure 4** is the curve diagram and **Figure 5** is the bar diagram. For the former, the X-axis demonstrates three kinds of training sizes and the Y-axis demonstrates values of MAE results. For the latter, only the two most typical training sizes are utilized for evaluation. Thus, it has two clusters of bars, corresponding to RMSE results under two training sizes. It is clearly observed that the GNN-APM is always endowed with better prediction efficiency compared with baseline methods. To sum up, this group of simulative experiments well demonstrates the good performance of the designed GNN-APM.

In order to visualize the tendency of MAE results and RMSE results under different experimental settings, some of the results are illustrated with the use of curve diagrams or bar diagrams. **Figure 4** illustrates the MAE results of three methods under two different learning rate values: 0.001 and 0.002. It has two subfigures that correspond to results about two learning rate values. In each subfigure, the X-axis denotes the proportion of training data ranging from 50 to 70%, and the Y-axis denotes values of MAE results. **Figure 5** illustrates RMSE results of three methods when the training data size is set to 60 and 70%. This is because it can be seen from previous experiments that result under the two training data sizes are relatively better. It has two subfigures that correspond to two learning rate values. In each subfigure, the X-axis denotes two training sizes, and the Y-axis denotes values of RMSE results. It can be observed from these figures that values of GNN-APM are obviously below the other two methods and that values show descending tendency when the proportion of training data increases. Such results demonstrate the improvement process of methods with being trained more sufficiently. These figures show a better performance tendency of GNN-APM compared with two other baseline methods.

5.3. Parameter Sensitivity

Besides, it is also expected to explore parameter sensitivity of the GNN-APM, and relevant simulative results are illustrated in **Figure 6**. During this group of experiments, the GNN-APM is not compared with baseline methods and just performance of itself is explored. **Figure 6** has two subfigures, corresponding to sensitivity results using two different metrics: MAE and RMSE. Inside each subfigure, the X-axis denotes the change of learning rate, and the Y-axis denotes the change of training size. In the middle square area, the color depth indicates the different values of evaluation metrics. As the two subfigures are heatmaps, the color depth degree inside figures is able to indicate values of metrics. Each subfigure includes a squared area, gentle color change inside it indicates that the performance of GNN-APM fluctuates not heavily. It can be objectively found that color fluctuation in both two subfigures seems quite gentle, revealing that the GNN-APM is not susceptible to parameter change. In other words, the GNN-APM is always able to remain stable, no matter how the key parameters change. This group of simulative results well prove that the GNN-APM possesses proper robustness.

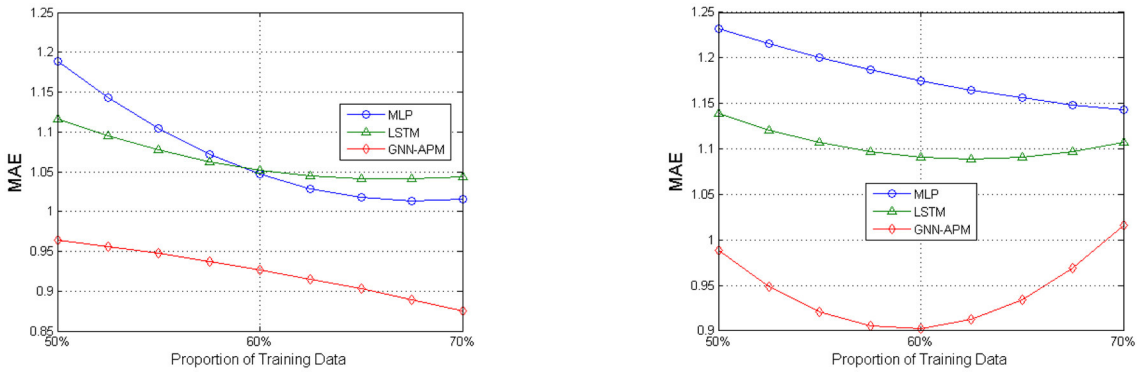


FIGURE 4 | Average mean absolute error (MAE) results under two different learning rate values.

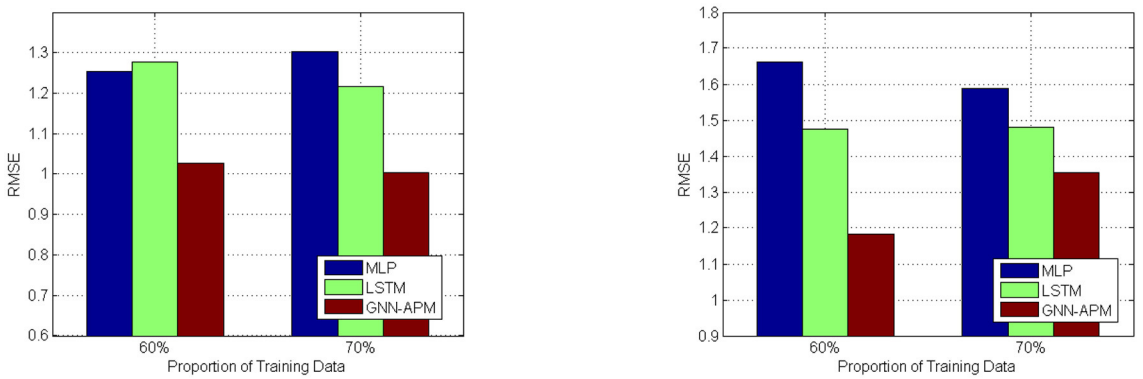


FIGURE 5 | Average RMSE results under two different learning rate values.

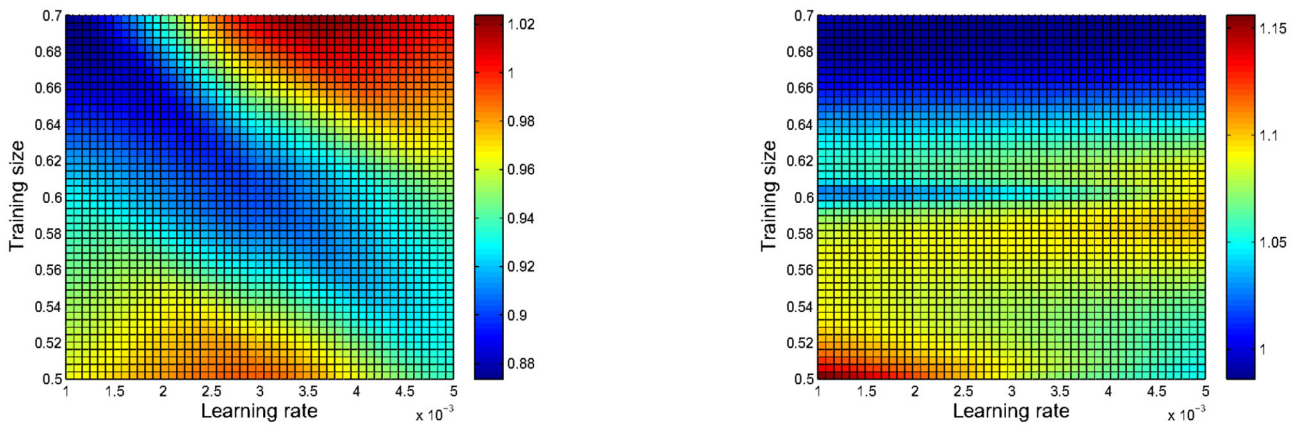


FIGURE 6 | Parameter sensitivity results of the designed GNN-APM concerning MAE and RMSE.

6. CONCLUSION

Agriculture has been viewed as the most fundamental industry since ancient times. Nowadays, E-commerce is an important sales channel of agricultural products. To better manage and schedule the supply of agricultural products, dynamic price prediction for agricultural products in the E-commerce market is of great significance. To overcome the shortcomings of

existing research studies, this article proposes a deep learning-based price prediction model for agricultural products in the E-commerce market. In particular, the most typical GCN is utilized to establish a time-series prediction model for the dynamic price of agricultural products. In addition, the whole simulative experiments are composed of three parts. First, the fluctuation tendency of the involved five agricultural product types is visualized using a curve diagram. Second, the prediction efficiency of the GNN-APM on five objects is compared with two baseline methods. Third, the robustness of the GNN-APM is tested by changing several parameter combinations.

Nowadays, data mining and data management for many industries are gradually approaching the application of the Internet of Things (IoT), yielding such as mobile IoT (36, 37), financial IoT, medical IoT (38), cloud-assisted IoT (39), vehicular IoT (40), and industrial IoT (41, 42). As is known to all, the IoT is a kind of effective tool or platform to integrate multi-domain data and schedule business flows. To realize more effective scheduling management of the agricultural product market, designing an integrated microservice IoT platform that is embedded with robust artificial intelligence algorithms (43), is in urgent demand to deal with many disturbing issues in various industries. Thus, for future outlook, the authors plan to deeply investigate optimal scheduling and management schemes for the agricultural product market with the use of novel IoT-related technologies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: http://www.moa.gov.cn/.

AUTHOR CONTRIBUTIONS

WY contributed significantly to theoretical analysis and manuscript preparation. ZZ performed the experiments and handled funding details. QZ contributed to the conception of the study and model formulation. GZ performed the data analyses and data visualization. QH provided many promising insights during the revision process. QL helped perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by Beijing Natural Science Foundation (No. 4202014), Natural Science Foundation of China (61873027), in part by Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 20YJCZH229), in part by the Open Project Program of National Engineering Laboratory for Agri-Product Quality Traceability (AQT-2020-YB8), and in part by Hubei Natural Science Foundation (No. 2021CFB156).

REFERENCES

1. Xiong H, Jin C, Alazab M, Yeh KH, Wang H, Gadekallu TRR, et al. On the design of blockchain-based ECDSA with fault-tolerant batch verication protocol for blockchain-enabled IoMT. *IEEE J Biomed Health Inform*. (2021). doi: 10.1109/JBHI.2021.3112693
2. Yu K, Tan L, Yang C, Choo KKR, Bashir AK, Rodrigues JJPC, et al. A blockchain-based shamir's threshold cryptography scheme for data protection in industrial internet of things settings. *IEEE Internet Things J*. (2021). doi: 10.1109/JIOT.2021.3125190
3. Wang W, Qiu C, Yin Z, Srivastava G, Gadekallu TR, Alsolami F, et al. Blockchain and PUF-based lightweight authentication protocol for wireless medical sensor networks. *IEEE Internet Things J*. (2021). doi: 10.1109/JIOT.2021.3117762
4. Yu K, Tan L, Shang X, Huang J, Srivastava G, Chatterjee P. Efficient and privacy-preserving medical research support platform against COVID-19: a blockchain-based approach. *IEEE Consumer Electron Mag*. (2021) 10:111–20. doi: 10.1109/MCE.2020.3035520
5. Zeng W, Guo Z, Yu K, Shen Y, Bashir AK, Al-Otaibi YD. Data-driven management for fuzzy sewage treatment processes using hybrid neural computing. *Neural Comput Appl*. (2021). doi: 10.1007/s00521-020-05655-3
6. Wang W, Fida MH, Lian Z, Yin Z, Pham QV, Gadekallu TR, et al. Secure-Enhanced federated learning for ai-empowered electric vehicle energy prediction. *IEEE Consumer Electron Mag*. (2021). doi: 10.1109/MCE.2021.3116917
7. Guo Z, Wang H. A deep graph neural network-based mechanism for social recommendations. *IEEE Trans Ind Inform*. (2021) 17:2776–83. doi: 10.1109/TII.2020.2986316
8. Guo Z, Yu K, Li Y, Srivastava G, Lin JCW. Deep learning-embedded social internet of things for ambiguity-aware social recommendations. *IEEE Trans Netw Sci Eng*. (2021). doi: 10.1109/TNSE.2021.3049262
9. Deepa N, Durai R, Senthil KN, Srinivasan K, Chang CY, Bashir AK. An efficient ensemble VTOPES multi-criteria decision-making model for sustainable sugarcane farms. *Sustainability*. (2019) 11:4288. doi: 10.3390/su11164288
10. Li H, Chen X, Guo Z, Xu J, Shen Y, Gao X. Data-driven peer-to-peer blockchain framework for water consumption management. *Peer-to-Peer Netw Appl*. (2021) 14:2887–900. doi: 10.1007/s12083-021-01121-6
11. Meng D, Xiao Y, Guo Z, Jolfaei A, Qin L, Lu X, et al. A data-driven intelligent planning model for UAVs routing networks in mobile internet of things. *Comput Commun*. (2021) 179:231–41. doi: 10.1016/j.comcom.2021.08.014
12. Elavarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: a survey. *Comput Electron Agric*. (2018) 155:24. doi: 10.1016/j.compag.2018.10.024
13. Ahmmadi P, Rahimian M, Movahed RG. Theory of planned behavior to predict consumer behavior in using products irrigated with purified wastewater in Iran consumer. *J Cleaner Product*. (2021) 296:126359. doi: 10.1016/j.jclepro.2021.126359
14. Fan P, Wang Y, Xu N. Value added mechanism and organisational model optimisation of agricultural products circulation value chain from the perspective of game theory. *Acta Agric Scand B Soil Plant Sci*. (2021) 71:215–23. doi: 10.1080/09064710.2021.1879927
15. Assa H, Sharifi H, Lyons A. An examination of the role of price insurance products in stimulating investment in agriculture supply chains for sustained productivity. *Eur J Oper Res*. (2021) 288:918–34. doi: 10.1016/j.ejor.2020.06.030
16. Ray P. Agricultural supply chain risk management under price and demand uncertainty. *Int J Syst Dyn Appl*. (2021) 10:17–32. doi: 10.4018/IJSDA.2021040102
17. Yang B, Chang M, Wang K, Zhang R. The calendar effect of price-reduction auction of online agricultural products. *Int J Serv Technol Manag*. (2021) 27:229–47. doi: 10.1504/IJSTM.2021.115167
18. Zhao H. Futures price prediction of agricultural products based on machine learning. *Neural Comput Appl*. (2021) 33:837–50. doi: 10.1007/s00521-020-05250-6

19. Zhang D, Chen S, Liwen L, Xia Q. Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons. *IEEE Access*. (2020) 8:28197–209. doi: 10.1109/ACCESS.2020.2971591
20. Huang J, Chen Y, He S. The evolutionary analysis of agricultural production transaction under the price subsidy policy. *Int J Inf Syst Supply Chain Manag*. (2020) 13:73–97. doi: 10.4018/IJISSCM.2020010104
21. Yadava TG, Jayanna HS. Improvements in spoken query system to access the agricultural commodity prices and weather information in Kannada language/dialects. *J Intell Syst*. (2020) 29:664–87. doi: 10.1515/jisys-2018-0120
22. Zheng G, Zhang H, Han J, Zhuang C, Xi L. The research on agricultural product price forecasting service based on combination model. In: *13th IEEE International Conference on Cloud Computing, CLOUD 2020 Virtual Event, 18-24 October 2020*. Beijing: IEEE (2020). p. 4–9.
23. Guo H, Woodruff A, Yadav A. Improving lives of indebted farmers using deep learning: predicting agricultural produce prices using convolutional neural networks. In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020 The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020 The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020 New York, NY, USA, February 7–12, 2020 AAAI Press*. New York, NY: AAAI Press (2020). p. 13294–9. doi: 10.1609/aaai.v34i08.7039
24. Chen L, Li Z. Research on the price analysis and prediction method of agricultural products based on logistics information. *Clust Comput*. (2019) 22:14951–57. doi: 10.1007/s10586-018-2462-y
25. Alizamir S, Iravani F, Mamani H. An analysis of price vs. Revenue protection: government subsidies in the agriculture industry. *Manag Sci*. (2019) 65:32–49. doi: 10.1287/mnsc.2017.2927
26. Madaan L, Sharma A, Khandelwal P, Goel S, Singla P, Seth A. Price forecasting & anomaly detection for agricultural commodities in India. In: *Proceedings of the Conference on Computing & Sustainable Societies, COMPASS 2019 Accra, Ghana, July 3–5 2019*. Accra: ACM (2019). p. 52–64.
27. Silva J, Gaitán-Angulo M, Borré JR, Ayarza LPL, Lezama OBP, del Carmen Martínez Galán Z, et al. An early warning method for agricultural products price spike based on artificial neural networks prediction. In: *Computational Science - ICCS 2019 - 19th International Conference, Faro, Portugal, June 12–14, 2019 Proceedings, Part II. Vol. 11537 of Lecture Notes in Computer Science*. Faro: Springer (2019). p. 622–32.
28. Deng Y, Zhu Y, Zhang Q, Cheng X. Hadoop + spark platform based on big data system design of agricultural product price analysis and prediction by holtwinters. In: *Data Science - 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2019 Guilin, China, September 20-23, 2019 Proceedings, Part I. Vol. 1058 of Communications in Computer and Information Science*. Springer (2019). p. 103–23.
29. Dharavath R, Khosla E. Seasonal ARIMA to forecast fruits and vegetable agricultural prices. In: *IEEE International Symposium on Smart Electronic Systems, iSES 2019 (Formerly iNiS), Rourkela, India, December 16-18, 2019*. Rourkela: IEEE (2019). p. 47–52.
30. Chen Q, Lin X, Zhong Y, Xie Z. Price prediction of agricultural products based on wavelet analysis-LSTM. In: *2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, ISPA/BDCloud/SocialCom/SustainCom*. Xiamen: IEEE (2019). p. 984–90.
31. Xiong T, Li C, Bao Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: evidence from the vegetable market in China. *Neurocomputing*. (2018) 275:2831–44. doi: 10.1016/j.neucom.2017.11.053
32. Kurumatani K. Time series prediction of agricultural products price based on time alignment of recurrent neural networks. In: Wani MA, Kantardzic MM, Mouchaweh MS, Gama J, Lughofer E, editors. *17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018 Orlando, FL, USA, December 17-20*. Orlando, FL: IEEE (2018). p. 81–8.
33. Guo Z, Yu K, Jolfaei A, Bashir AK, Almagrabi AO, Kumar N. A fuzzy detection system for rumors through explainable adaptive learning. *IEEE Trans Fuzzy Syst*. (2021) 29:3650–64. doi: 10.1109/TFUZZ.2021.3052109
34. Yu K, Tan L, Lin L, Chen X, Yi Z, Sato T. Deep-Learning-Empowered breast cancer auxiliary diagnosis for 5GB remote e-health. *IEEE Wirel Commun*. (2021) 28:54–61. doi: 10.1109/MWC.001.2000374
35. Guo Z, Shen Y, Bashir AK, Yu K, wei Lin JC. Graph embedding-based intelligent industrial decision for complex sewage treatment processes. *Int J Intel Syst*. (2021). doi: 10.1002/int.22540
36. Zhang Q, Yu K, Guo Z, Garg S, Rodrigues JJPC, Hassan MM, et al. Graph neural networks-driven traffic forecasting for connected internet of vehicles. *IEEE Trans Netw Sci Eng*. (2021). doi: 10.1109/TNSE.2021.3126830
37. Yu K, Guo Z, Shen Y, Wang W, Lin JCW, Sato T. Secure artificial intelligence of things for implicit group recommendations. *IEEE Internet Things J*. (2021) 9:2698–707. doi: 10.1109/JIOT.2021.3079574
38. Sun Y, Liu J, Yu K, Alazab M, Lin K. PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare. *IEEE Trans Ind Inform*. (2021) 18:1981–90. doi: 10.1109/TII.2021.3070544
39. Wang T, Quan Y, Shen XS, Gadekallu TR, Wang W, Dev K. A privacy-enhanced retrieval technology for the cloud-assisted internet of things. *IEEE Trans Ind Inform*. (2021). doi: 10.1109/TII.2021.3103547
40. Feng C, Yu K, Aloqaily M, Alazab M, Lv Z, Mumtaz S. Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV. *IEEE Trans Veh Technol*. (2020) 69:13784–95. doi: 10.1109/TVT.2020.3027568
41. Li H, Yu K, Liu B, Feng C, Qin Z, Srivastava G. An efficient ciphertext-policy weighted attribute-based encryption for the internet of health things. *IEEE J Biomed Health Inform*. (2021). doi: 10.1109/JBHI.2021.3075995
42. Wang W, Xu H, Alazab M, Gadekallu TR, Han Z, Su C. Blockchain-Based reliable and efficient certificateless signature for IIoT devices. *IEEE Trans Ind Inform*. (2021). doi: 10.1109/TII.2021.3084753
43. Tan L, Yu K, Shi N, Yang C, Wei W, Lu H. Towards secure and privacy-preserving data sharing for COVID-19 medical records: a blockchain-empowered approach. *IEEE Trans Netw Sci Eng*. (2021) 9:271–81. doi: 10.1109/TNSE.2021.3101842

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yan, Zhang, Zhang, Zhang, Hua and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fpubh.2022.879639

Medical Image Segmentation Algorithm for Three-Dimensional Multimodal Using Deep Reinforcement Learning and Big Data Analytics

Weiwei Gao[1], *Xiaofeng Li*[2]*, *Yanwei Wang*[3] *and Yingjie Cai*[4]

[1] College of Information and Technology, Wenzhou Business College, Wenzhou, China, [2] Department of Information Engineering, Heilongjiang International University, Harbin, China, [3] School of Mechanical Engineering, Harbin Institute of Petroleum, Harbin, China, [4] The First Psychiatric Hospital of Harbin, Harbin, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Haitao Yang,
Beijing University of Technology, China
Cheng Wang,
Northwestern Polytechnical University, China

***Correspondence:**
Xiaofeng Li
lixiaofeng@hiu.net.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 20 February 2022
Accepted: 09 March 2022
Published: 08 April 2022

Citation:
Gao W, Li X, Wang Y and Cai Y (2022) Medical Image Segmentation Algorithm for Three-Dimensional Multimodal Using Deep Reinforcement Learning and Big Data Analytics. Front. Public Health 10:879639. doi: 10.3389/fpubh.2022.879639

To avoid the problems of relative overlap and low signal-to-noise ratio (SNR) of segmented three-dimensional (3D) multimodal medical images, which limit the effect of medical image diagnosis, a 3D multimodal medical image segmentation algorithm using reinforcement learning and big data analytics is proposed. Bayesian maximum a posteriori estimation method and improved wavelet threshold function are used to design wavelet shrinkage algorithm to remove high-frequency signal component noise in wavelet domain. The low-frequency signal component is processed by bilateral filtering and the inverse wavelet transform is used to denoise the 3D multimodal medical image. An end-to-end DRD U-Net model based on deep reinforcement learning is constructed. The feature extraction capacity of denoised image segmentation is increased by changing the convolution layer in the traditional reinforcement learning model to the residual module and introducing the multiscale context feature extraction module. The 3D multimodal medical image segmentation is done using the reward and punishment mechanism in the deep learning reinforcement algorithm. In order to verify the effectiveness of 3D multimodal medical image segmentation algorithm, the LIDC-IDRI data set, the SCR data set, and the DeepLesion data set are selected as the experimental data set of this article. The results demonstrate that the algorithm's segmentation effect is effective. When the number of iterations is increased to 250, the structural similarity reaches 98%, the SNR is always maintained between 55 and 60 dB, the training loss is modest, relative overlap and accuracy all exceed 95%, and the overall segmentation performance is superior. Readers will understand how deep reinforcement learning and big data analytics test the effectiveness of 3D multimodal medical image segmentation algorithm.

Keywords: deep reinforcement learning, three-dimensional multimodal, wavelet shrinkage, medical image segmentation, high-frequency signal component

INTRODUCTION

In current medical practice, medical images such as MRI image, CT image, and ultrasonic imaging are important diagnostic basis for diagnosing patients' physical conditions. Doctors or researchers analyze the tissues and organs in the human body through the medical image, obtain the relevant information of the tissues and organs in the human body according to the medical image, and carry out treatment planning in combination with medical knowledge. Therefore, medical image has become an indispensable part of disease diagnosis and treatment and quality. Medical image segmentation is not only a prerequisite for computer-aided diagnosis and many medical image applications, but also an important stage of medical image visual analysis (1, 2). Image segmentation of medical image can help to directly obtain the contour of the target by identifying the target of interest. In diagnosis, it is convenient for doctors to obtain the patient's condition information (3). With the development of medicine, medical images have been continuously optimized to gradually form three-dimensional (3D) multimodal medical images, which make medical images clearer and have higher resolution (4, 5). In order to effectively distinguish the pathological region from the normal region in medical image and enable doctors to diagnose and treat more intuitively, the segmentation of 3D multimodal medical image has become the focus of current research. However, in practical application, the factors such as image device, organ position, and shape still hinder the high-precision segmentation of medical image. The traditional image segmentation algorithm is difficult to overcome the influence of many factors and the effect of segmented medical image is poor.

At present, with the continuous improvement of science and technology, deep reinforcement learning technology has been applied in a large number of fields because this technology has high feature extraction ability. Compared with the traditional segmentation algorithm, it has obvious advantages and is widely used in the process of image segmentation (6). Therefore, a 3D multimodal medical image segmentation algorithm using deep reinforcement learning is proposed in this article. The image segmentation is performed by constructing a deep reinforcement learning model. The main contributions of this article are as follows: (1) We combine wavelet denoising with bilateral filtering to realize 3D multimodal medical image denoising. According to the traditional wavelet denoising method, Bayesian maximum a posteriori estimation method is used to realize wavelet shrinkage and combined with bilateral filter to improve the denoising performance of image; (2) The extended convolution is added and two-dimensional convolution is introduced to improve the receptive field and ensure that comprehensive context information can be obtained during segmentation; and (3) In deep reinforcement learning, the residual network is introduced to make the segmentation results more accurate and improve the training speed of deep reinforcement learning. Using the introduced residual structure, the feature extraction performance of the network can be improved.

RELATED WORKS

Many scholars have studied the related problems of medical image segmentation. Literature (7) studied medical image segmentation based on spatial constraints and fuzzy spatial segmentation. By analyzing the image edge position, the realization of spatial constraints and complete medical image segmentation were determined. However, the algorithm is only suitable for fuzzy and uncertain medical image processing. When it is used for 3D multimodal medical image, the segmentation performance of the algorithm is poor. Literature (8) studied the weakly supervised retinal vascular segmentation algorithm and used the hierarchical clustering algorithm to classify the vascular and non-vascular pixels. Moreover, the classification results based on the driving database were used as the basic facts to train the neural network. By calculating and comparing the image feature differences between the target domain data and the source domain data extracted from the network, the data required for training were extended based on semi-supervised clustering to realize retinal vessel segmentation. The algorithm can make pixels highly reliable, but the algorithm can improve the segmentation accuracy of retinal vessels, but its segmentation effect is poor. Literature (9) studied the automatic segmentation of lung tumors in CT images based on multiresolution residual connected feature flow. Incremental and dense multiresolution residual connected networks were used to detect and segment lung tumors by combining the features of multiple image resolutions and feature levels. The algorithm can effectively improve the image dimension in the segmentation process, but the segmentation operation can only be realized according to the feature flow, so the segmentation quality factor is poor. Literature (10) studied image segmentation combining image weighting and transfer learning and used kernel learning as a method to reduce the difference between training and test data, so as to enhance the performance of image segmentation through continuous learning and training. However, the algorithm does not consider the multimodality of 3D image in the segmentation process, so the definition is poor. Literature (11) studied Pulse Coupled Neural Networks (PCNN) medical image segmentation based on multifeature gray wolf optimized bionic algorithm and used image multifeatures to realize segmentation. However, the algorithm has poor structural similarity between training and actual segmentation, cannot completely realize segmentation, and the signal-to-noise ratio (SNR) is low, resulting in more noise points in the image. Literature (12) studied medical image segmentation based on dice score and the Jaccard index and the relationship within the measurement sensitive loss function group from a theoretical point of view. Moreover, the existence of the optimal weighting scheme of weighted cross entropy was questioned to optimize the dice score and the Jaccard index during the test and realize medical image segmentation through dice score and the Jaccard index. This method optimized medical image segmentation to some extent, but it did not denoise the image, so the effect of medical image segmentation is poor. Literature (13) studied medical image segmentation based on context feedback loop and expressed the segmentation

problem as a recursive framework by using two systems. The first is the forward system of encoder-decoder convolutional neural network, which predicts the segmentation results from the input image. The predicted forward system probability output is encoded by a context feedback system based on a complete convolution network. Then, the coding feature space of the complete convolution network is integrated back into the feedforward learning process of the forward system. This method uses the context feedback loop based on complete convolution network and the forward system can learn and extract more advanced image features. Moreover, it fixes the previous errors and improves the segmentation accuracy over time. This method can effectively improve the segmentation accuracy, but it does not denoise the collected medical image, so that the segmentation quality factor is low.

In the process of image segmentation by the above methods, the analysis of influencing factors in the process of image segmentation is not comprehensive, resulting in low relative overlap and signal-to-noise ratio of image segmentation 3D multimodal medical image. Therefore, this article studies the 3D multimodal medical image segmentation algorithm based on deep reinforcement learning and uses wavelet algorithm and bilateral filter to denoise the image to make the image clearer. Then, a deep reinforcement learning model is constructed to realize multimodal medical image segmentation by combining residual structure and dilated residual and deeply supervised U-Net (DRD U-Net) convolution model. The performance of the proposed algorithm is verified by experiments.

METHODOLOGY

Medical Image Denoising of Three-Dimensional Multimodal Based on Wavelet and Bilateral Filtering

Combined with wavelet denoising and bilateral filtering, 3D multimodal medical image denoising is realized. According to the traditional wavelet denoising method, Bayesian maximum a posteriori estimation method is used to realize wavelet shrinkage and combined with bilateral filter to improve the denoising performance of image.

Setting of Wavelet Threshold Function

In order to realize the accurate segmentation of medical image segmentation, it is necessary to denoise the image. In this article, the wavelet threshold function is set to provide the prerequisite for image denoising. The calculation formula of wavelet shrinkage threshold function is constructed $T = \sigma_N\sqrt{2\log M}$ for analysis.

where M refers to the total number of wavelet coefficients in the corresponding wavelet domain. Because the calculation effect of the general threshold function is not perfect, the general threshold function is redesigned in an improved form, as shown in Equation (1).

$$T_j = a_j\sigma_N\sqrt{2\log M} \tag{1}$$

where after decomposition, the number of layers with wavelet coefficients is described by $j = (1, 2, ..., J)$. The maximum number of decomposition layers is J, a_j is the adaptive parameter, and the adaptive parameter corresponding to layer j is 2^{J-j+1}.

Design of Wavelet Shrinkage Algorithm

In order to obtain the threshold function of wavelet shrinkage, according to the constructed wavelet shrinkage threshold function, the wavelet coefficients of noiseless signal are analyzed by generalized Laplace distribution. Moreover, the Bayesian maximum a posteriori estimation method is used to calculate the a posteriori probability and complete the design of wavelet shrinkage algorithm.

Wavelet coefficients of noiseless signals can be analyzed by generalized Laplace distribution $G^j_{l,k}$ and calculate its probability distribution through Equation (2).

$$p_G(g) = \frac{v}{2s\Gamma\left(\frac{1}{v}\right)}\exp\left(-\left|\frac{g}{s}\right|^v\right), s, v > 0 \tag{2}$$

where $p_G(g)$ refers to the probability distribution of the coefficient, g refers to noise-free signal, subscript (l, k) is coordinates in wavelet domain, the gamma function is $\Gamma(a) = \int_0^\infty x^{a-1}\exp(-x)\,dx$, and the scale parameter is s. If shape parameter $v = 1$, Equation (2) can be changed to Laplace distribution, as shown in Equation (3).

$$p_G(g) = \frac{1}{2s}\exp\left(-\left|\frac{g}{s}\right|\right), s > 0 \tag{3}$$

The speckle noise $N^j_{l,k}$ in the wavelet domain can be calculated by Gaussian distribution, as shown in Equation (4).

$$p_N(n) = \frac{1}{\sqrt{2\pi}\sigma_N}\exp\left(-\frac{n^2}{2\sigma_N^2}\right) \tag{4}$$

where $p_N(n)$ refers to the probability distribution of the speckle noise and the noise SD in wavelet domain is described by σ_N. In order to obtain the signal estimation in the domain, the Bayesian maximum a posteriori estimation method is used to calculate the a posteriori probability. It is obtained by Equation (5).

$$\begin{aligned} &p_{G|F}(g|f) = \frac{1}{p_F(f)}p_{F|G}(f|g)\,p_G(g) = \\ &\frac{1}{p_F(f)}p_N(f-g)\,p_G(g) \end{aligned} \tag{5}$$

where f refers to the medical image network signal obtained in the initial state and Equation (3) and Equation (4) are introduced into Equation (5) to obtain Equation (6).

$$\begin{aligned} &p_{G|F}(g|f) = \frac{1}{p_F(f)}\cdot\frac{1}{2\sqrt{2\pi}s\sigma_N}\times \\ &\exp\left(\frac{2\sigma_n^2|g| - s(f-g)^2}{2s\sigma_N^2}\right) \end{aligned} \tag{6}$$

To calculate the maximum a posteriori probability, set the result of derivation calculation from $\ln\left(p_{G|F}(g|f)\right)$ to g is 0. It is obtained by Equation (7).

$$\hat{g} = \mathrm{sgn}(f)\max\left(|f| - \frac{\sigma_N^2}{s}, 0\right) \tag{7}$$

where g is estimated to be $\hat{g}$ and set f and g are of the same number. The threshold function of wavelet shrinkage is obtained through the Equation (7) calculation. Therefore, the wavelet shrinkage algorithm is expressed by Equation (8).

$$\hat{g} = \begin{cases} 0 & f \le T_j \\ \operatorname{sgn}(f) \max\left(|f| - \frac{\sigma_N^2}{s}, 0\right) & f > T_j \end{cases} \tag{8}$$

Combination With Bilateral Filter

The traditional wavelet denoising methods retain the unchanged wavelet coefficients in the low-frequency domain and only deal with the threshold of wavelet coefficients in the high-frequency domain, so the denoising effect is not obvious (14, 15). In order to remove the noise spots in the low-frequency domain, this article combines the bilateral filter to filter the wavelet coefficients in the low-frequency domain. The structure of the bilateral filter is represented by Equation (9).

$$h(x) = k^{-1}(x) \times \int_{\xi \in \Omega(x)} f(\xi)\, c(\xi, x)\, s\left(f(\xi), f(x)\right) d\xi \tag{9}$$

where the normalization factor is described by $k(x) = \int_{\xi \in \Omega(x)} c(\xi, x)\, s\left(f(\xi), f(x)\right) d\xi$ and the window region with the pixel x as the center is $\Omega(x)$. A bilateral filter is formed by combining two filter cores (16, 17), i.e., the regional kernel $c(\xi, x)$ and range kernel $s\left(f(\xi), f(x)\right)$. In the above formula, the distance function from the edge pixel ξ in region $\Omega(x)$ and the pixel of the central region x is $c(\xi, x)$. Meanwhile, in the $\Omega(x)$ region, the similarity function between the pixel value cc of the edge pixel point ξ and the pixel value $f(\xi)$ of the pixel point $f(x)$ in the center region is described by $s\left(f(\xi), f(x)\right)$. After bifiltering by $f(x)$, the calculation result is shown by $h(x)$.

Steps of Medical Image Denoising

In order to obtain the denoised medical image, the ultrasonic network signal collected by the ultrasonic imaging system needs to be processed by logarithmic transformation.

(1) Log transforms the ultrasonic network signal collected by the ultrasonic imaging system. If 3D multimodal medical image is directly collected, this step is not required.

(2) Through the image obtained in wavelet decomposition processing step (1), four frequency domains are obtained, which are LL^1, LH^1, HL^1, and HH^1. Continue wavelet decomposition of LL^1 in low-frequency domain and obtain four frequency domains again (18, 19), in order: LL^2, LL^2, HL^2, and HH^2. Perform wavelet decomposition repeatedly until the maximum number of layers is decomposed J.

(3) The bottom low-frequency region LL^J is processed by bilateral filtering of Equation (9).

(4) The wavelet coefficients in the high-frequency region (LH^j, HL^j, HH^j, $j = 1, 2, ..., J$) of each layer are shrunk by Equation (8) and the SD of noise, image, and noise-free signal in each frequency domain is calculated, respectively.

(5) The inverse wavelet transform is performed on the above parameters to obtain the denoised medical image and complete the denoising.

DRD U-Net Model Using Deep Reinforcement Learning

This article selects the end-to-end DRD U-Net model to segment the 3D multimodal medical image after denoising. The convolution part of the traditional reinforcement learning model is adjusted to the residual module (8, 20) and multiscale context feature extraction atrous spatial pyramid pooling (ASPP) is introduced into the model, so that images of different sizes can be segmented.

Multiscale Context Feature Extraction Module

In order not to increase the parameters rapidly, this article adds extended convolution and introduces two-dimensional convolution to improve the receptive field.

By expanding convolution, the receptive field can be increased when the parameters remain unchanged, so as to ensure that comprehensive context information can be obtained during segmentation (21). It is assumed that ASPP consists of four parallel extended convolutions. At the same time, set the module input as $x \in R^{4c \times h \times w}$, the expansion rate is set as [3, 5, 7, 9]. Then, Equation (10) can be used to calculate the output of the module.

$$m = concatenate\,(m_1, m_2, m_3, m_4) \tag{10}$$

where $m_i \in R^{c \times h \times w}$, according to the expansion convolution with different expansion rates, a large number of context features are obtained and then a variety of features are spliced to comprehensively extract image multiscale features.

Medical Image Segmentation Algorithm of DRD U-Net Model

The model introduces residual network in deep reinforcement learning to make the segmentation result more accurate and improve the training speed of deep reinforcement learning.

Residual Structure

The residual module can improve the optimization performance of neural network, make the training process easier, and slow down the degradation speed of deep network. In this article, batch normalization (BN) operation is combined in the residual module and ReLu is selected as the excitation function. At the same time, in order to improve the problem of channel number mismatch during input and output, 1×1 convolution is used to speed up the channel number transformation and a large number of parameters are not added to the convolution.

The residual structure of the module actually refers to the introduction of a bypass, so that the input and output data can be added to obtain the deep reinforcement learning model. The introduced bypass is usually called shortcut. Using the residual structure, the convergence speed of learning can be increased. Suppose b_1 is network input, then the output of a residual network unit can be represented by Equation (11).

$$m_1 = h\left(b_1\right) + I\left(b_1, w_1\right) \tag{11}$$

where the function to be fitted in the network is described by $h(\cdot)$, the convolution calculation is described by $I\left(b_1, w_1\right)$, the weight parameter in the network is w_1, and the first layer network output

is m_1. If $h(\cdot)$ belongs to identity mapping, $h(b_1) = b_1$. Therefore, the network output m_1 of the first floor residual structure can be calculated by Equation (12).

$$m_1 = b_1 + 1 = b_1 + I(b_1, w_1) \tag{12}$$

After the initial data of the input image and the output data after convolution calculation are linearly added, it is all the outputs of the residual unit. Therefore, one addition algorithm can be used to describe all the network outputs. If there are N residual structure network modules in the whole network, the overall network output can be calculated by Equation (13).

$$b_N = b_n + \sum_{i=n}^{N-1} F(b_i, W_i) \tag{13}$$

Medical Image Segmentation of Joint Residual Structure and DRD U-Net Convolution

For 3D multimodal medical image, the edge of medical image has high complexity, so it is usually difficult to effectively segment the lesion area from the normal area. Therefore, by adding the residual structure, the initial output characteristics of the image are retained to prevent the edge data around the tissue from being damaged during segmentation. At the same time, the local data in multimodal medical image can be fully combined during training.

In this article, the convolution layer in the residual module is set as 3D convolution form, a multiscale context feature extraction module is added, and the BN layer is adjusted to GN layer. According to the convolution kernel size, construct the convolution layer of $3 \times 3 \times 3$. Then, it is introduced into the GN layer to improve the convergence ability of the network and then input to the non-linear activation function layer. Next, it is input to a $3 \times 3 \times 3$ convolution kernel again. In order to accurately extract context features, two 3D convolution layers are used to increase the depth of the network. On this basis, the computational complexity can be effectively improved. Using the introduced residual structure, the feature extraction performance of the network can be improved.

Medical Image Segmentation Algorithm for Three-Dimensional Multimodal

With its powerful feature representation ability, deep reinforcement learning can accurately segment multidimensional and multimodal medical images. The 3D multimodal medical image segmentation algorithm based on deep reinforcement learning is as follows:

$$W = \frac{B}{H} \times \frac{u \cap v}{u \cup v} \tag{14}$$

where B refers to obtain the pixel value of the target of interest in the medical image, H refers to the total pixel value in the medical image obtained, v refers to the region where interested target locates, and v refers to the region where interested denoised target locates.

Negative rewards often occur in the process of deep reinforcement learning. In order to solve these problems, the reward method from the starting point of intermediate difficulty is used to calculate from the intermediate process. The reward calculation of medical image segmentation is as follows:

$$f(l_t, m_t) = sign(W(s_{t+1}, n) - W(s_t, n)) \tag{15}$$

where after the best segmentation m_t, get the next monitoring region s_{t+1}, and predict the reward according to the segmented W. If the image segmentation at time $t+1$ is greater than the image segmentation at time t, a reward is given. Otherwise, a penalty is given. Through the reward and penalty mechanism, the incorrect segmentation of medical images is minimized. The image segmentation at time t can be expressed by the following Equation:

$$Q = \sum_{t=1}^{t_{\max}} \kappa^{t-1} f_t \tag{16}$$

where $\kappa \in [0, 1]$ represents the discount factor. The greater its value, the greater the total reward.

To prevent local minima, use χ to mean the accurate qualitative strategy, χ begins from 0.9 and decreases by 0.1 each time till χ remains at 0.1.

$$m_t = \begin{cases} \text{Re-segmentation}, & \text{Under } \chi \text{ probability} \\ P(l_t, m_t), & \textit{Others} \end{cases} \tag{17}$$

Through the above calculation process, the 3D multimodal medical image segmentation is realized and the calculation process is shown in **Figure 1**.

EXPERIMENTAL RESULTS AND ANALYSIS

Data Set

In the experiment, PyTorch is used to build the structural framework of deep reinforcement learning and python language is used to complete the programming of medical image segmentation algorithm.

In order to verify the effectiveness of 3D multimodal medical image segmentation algorithm based on deep reinforcement learning, the algorithm is verified by experiments. The LIDC-IDRI data set, the Segmentation in Chest Radiographs (SCR) data set, and the DeepLesion data set are selected as the experimental data set of this article. The LIDC-IDRI data set: This dataset includes diagnostic and screening chest CT scans for lung cancer. A total of 1,018 study examples are included in the data set, with annotated lesions indicated. The SCR data set: This data set is a public database containing 247 chest images with segmentation of each image. The DeepLesion data set: This data set is an open dataset of CT images, including 32,735 CT images and lesion information from 4,427 patients, of which brain images and lung images are selected for this experiment. From the above three datasets, 5,000 medical images were selected for experimental analysis. The image sizes in the medical data were adjusted to 84×84 and the resolution of the images was $0.84 \times 0.84 \times 3$ mm. These experimental data were randomly sorted and 4,000 images were selected as the training data set and 1,000 images were selected as the test data set for the experiments in this article.

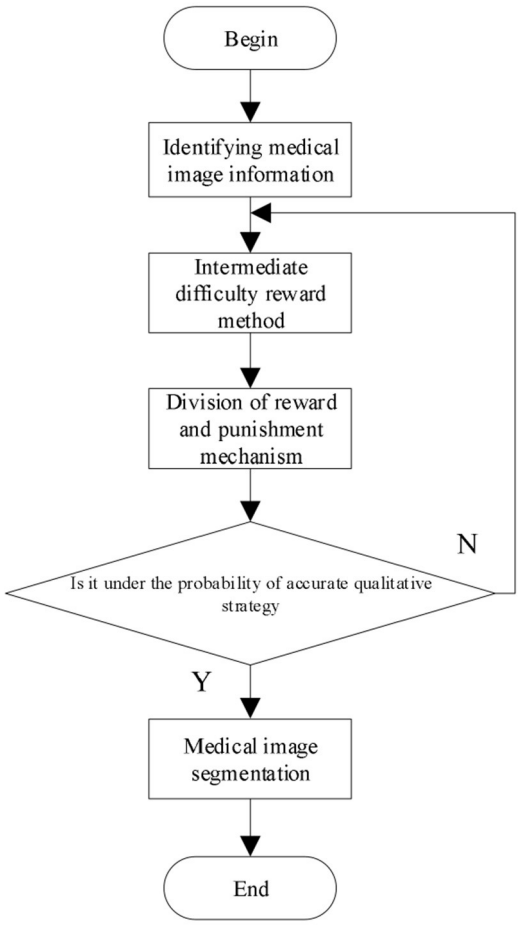


FIGURE 1 | Process of three-dimensional (3D) multimodal medical image segmentation.

Evaluation Criteria

1 Structural similarity: Medical image has a good structure and the pixels of image have many important correlations. The calculation procedure is shown in Equation (18).

$$SSIM\,(X,Y) = \left[l\,(X,Y)\right]^{\alpha}\left[c\,(X,Y)\right]^{\beta}\left[s\,(X,Y)\right]^{\gamma} \tag{18}$$

where brightness comparison is described by $l(X,Y)$, contrast comparison is described by $c(X,Y)$, image structure comparison is described by $s(X,Y)$, through parameters α, β, and γ, adjust the relative importance of the above three components. The greater the structural similarity, the better the segmentation effect.

2 Signal-to-noise ratio: The ratio between image information and noise is the SNR. Through the analysis of SNR, we can get the noise situation after image segmentation. When the SNR is higher, it means that the segmentation effect is better.

3 Training loss: The loss is a numerical value indicating the accuracy of image segmentation. If the segmented image is the same as the actual image, the loss is zero.

4 Relative overlap: The segmentation performance verification criteria is defined by the Equation (19) and the relative overlap is shown in Equation (19).

$$R_overlap = \frac{A\,(T \cap S)}{A\,(T \cup S)} \tag{19}$$

where the overall target area is described by $A\,(T)$, the size of the segmented target area is described by $A\,(S)$, and the higher the *R_overlap*, the more accurate the segmentation.

5 Accuracy: Accuracy analysis is performed on the test data set, when the greater the accuracy, the more accurate the segmented image is, using the Equation (20) to calculate the accuracy.

$$P = \frac{A\,(T \cap S)}{A\,(S)} \tag{20}$$

RESULTS AND DISCUSSION

In order to verify the effectiveness of the 3D multimodal medical image segmentation algorithm based on deep reinforcement learning, the proposed algorithm is validated by simulation experiments.

The proposed algorithm is used to segment the 3D multimodal medical images in the test set. Five brain images in the DeepLesion data set and three lung images in the SCR data set are selected for this study and the comparison of the effect pre- and postsegmentation is shown in **Figures 2**, **3**.

According to **Figures 2**, **3**, the application of the segmentation proposed algorithm can effectively achieve medical image segmentation. The image edge is clear for postsegmentation, the noise points can be effectively removed, and the definition is high.

In order to further verify the performance of the segmentation algorithm, the structural similarity, SNR, training loss, relative overlap, and accuracy are selected to verify the actual performance of the algorithm.

The training data set was trained with different iterations to analyze the structural similarity of the different algorithms after training and the analysis results are shown in **Figure 4**.

According to **Figure 4**, with the increase of the number of iterations, the structural similarity of different algorithms increases, which shows that the training of each algorithm can optimize the algorithm. Among them, the structural similarity of the algorithm in literature (11) remains the lowest in the last iteration. Therefore, the segmentation performance of the algorithm is poor, while the structural similarity of the algorithm in literature (8) increases rapidly during the iteration and finally reaches 96%. This shows that the algorithm can effectively ensure the structural similarity of the image during segmentation, but the algorithm is still lower than the proposed algorithm. When the number of training iterations is 250, the structural similarity reaches 98% and each iteration has the highest structural similarity. Therefore, compared with other algorithms, the segmentation training effect of proposed algorithm is better. The proposed algorithm combined with bilateral filter to filter the low-frequency wavelet coefficients and uses deep reinforcement

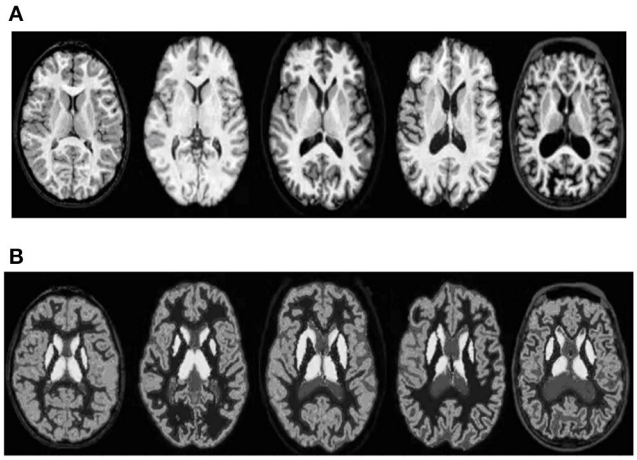


FIGURE 2 | Comparison of the effect of brain image segmentation before and after. **(A)** Presegmentation brain images. **(B)** Brain images postsegmentation by proposed algorithm.

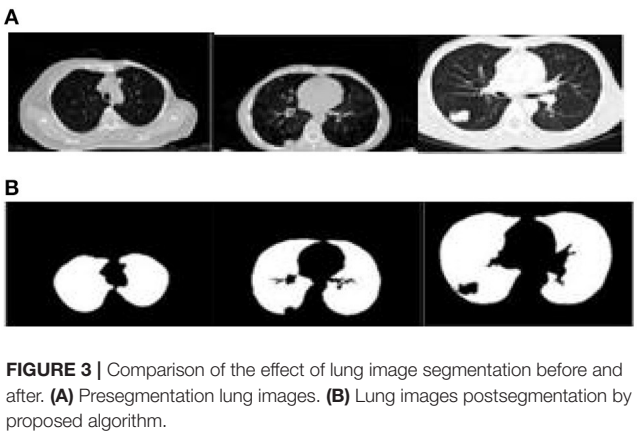


FIGURE 3 | Comparison of the effect of lung image segmentation before and after. **(A)** Presegmentation lung images. **(B)** Lung images postsegmentation by proposed algorithm.

learning to segment the medical image, which can enhance the structural similarity of image.

The ratio between image information and noise is the SNR. Through the analysis of SNR, the noise content after image segmentation can be obtained. The higher the SNR, the better the segmentation effect, so as to analyze the changes of the SNR of different algorithms after training. The results of the analysis are shown in **Figure 5**.

It is seen from **Figure 5** that when the number of iterations increases gradually, the SNR increases postsegmentation by different algorithms. Among them, the SNR of the algorithm in literature (8) increases greatly, gradually from 24 to 47 dB, but the SNR of the algorithm is still lower than that in literature (11) and the proposed algorithm. The SNR of the algorithm in literature (10) remains the lowest during iteration and there is no significant increase and the maximum SNR is only 33 dB, which shows that the algorithm has low noise reduction ability during training. The proposed algorithm always has a high SNR during training and is very stable. Under different iterations, the

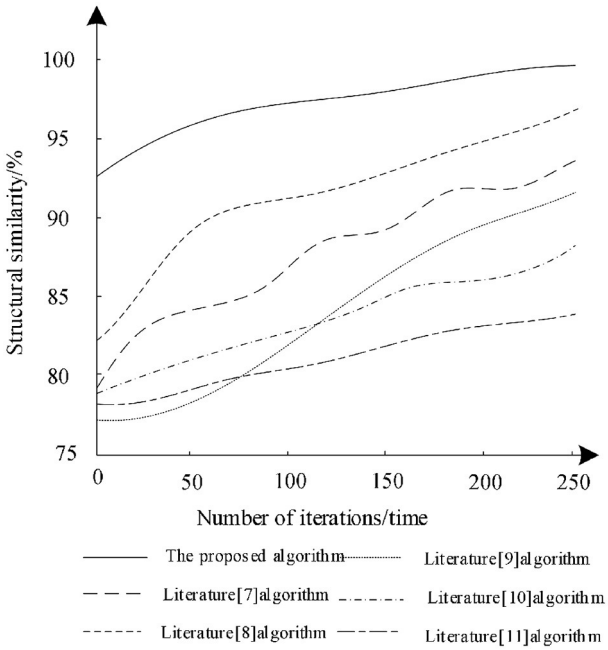


FIGURE 4 | Comparison of structural similarity of different algorithms.

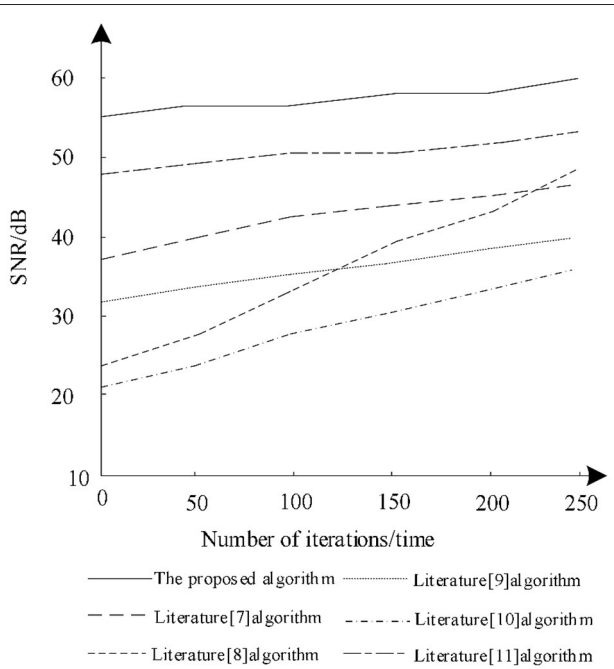


FIGURE 5 | Comparison of the signal-to-noise ratio (SNR) variation of different algorithms.

SNR is always maintained between 55 and 60 dB. Therefore, using proposed algorithm for segmentation can effectively reduce image noise.

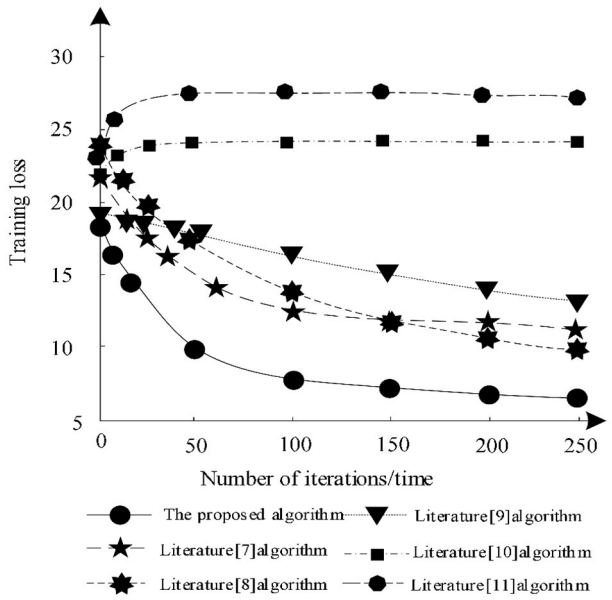


FIGURE 6 | Training loss of different algorithms.

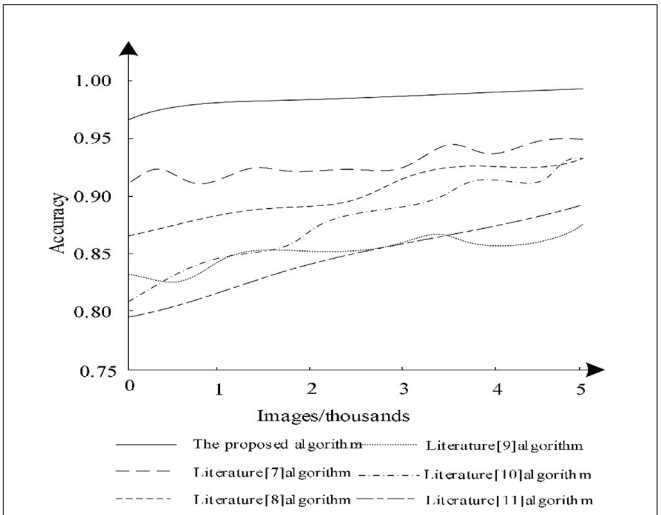


FIGURE 7 | Comparison of segmentation accuracy of different algorithms.

TABLE 1 | Comparison of the relative overlap of different algorithms.

Algorithms	Relative overlap		
	LIDC-IDRI	SCR	DeepLesion
The proposed	0.95	0.96	0.95
Literature (7)	0.89	0.80	0.84
Literature (8)	0.85	0.81	0.85
Literature (9)	0.87	0.81	0.81
Literature (10)	0.83	0.80	0.80
Literature (11)	0.85	0.84	0.87

The training loss of each algorithm segmentation under different training times is analyzed and the analysis results are shown in **Figure 6**.

According to **Figure 6**, when the number of iterations increases, the training losses of algorithm in literature (7), algorithm in literature (8), algorithm in literature (9), and algorithm in this article gradually decrease, while the training losses of algorithm in literature (10) and algorithm in literature (11) show an upward trend. Among them, the training loss of the algorithm in literature (11) is always the highest, indicating that the algorithm always has a large loss during segmentation training and cannot guarantee the integrity of the image. The training loss of the proposed algorithm decreases gradually in the iterative process, which shows that after training, the proposed algorithm can make the image segmentation more complete and ensure the effectiveness of segmentation. Because proposed algorithm uses DRD U-Net model, the model introduces residual network in deep reinforcement learning, which makes the segmentation result more accurate and the training loss less.

The relative overlap analysis is performed on the test image data set by comparing the proposed algorithm with those of literature (7), literature (8), literature (9), literature (10), and literature (11). The experimental results are shown in **Table 1**.

As can be seen from **Table 1**, the relative overlap of each algorithm reaches more than 80%. Among them, there are small differences in the relative overlap of the algorithms under the three data set tests of the LIDC-IDRI, the SCR, and the DeepLesion, but in terms of numerical size comparison, the proposed algorithm always maintains a high relative overlap of up to 95%, which is much higher than other algorithms in the literature, indicating that the proposed algorithm can achieve a more complete segmentation. Because the proposed algorithm uses deep reinforcement learning methods, with its powerful feature representation capability, it can accurately segment multidimensional multimodal medical images through its reward and punishment mechanism, thus improving the relative overlap of segmentation results.

The accuracy analysis was performed on the test dataset, when the larger the accuracy value, the better the image segmentation effect. The test images were analyzed using Equation (20) and the segmentation accuracy capabilities of different algorithms were compared and the analysis results are shown in **Figure 7**.

According to **Figure 7**, it can be seen that when the number of images increases, the segmentation accuracy of different algorithms changes. Among them, the algorithm of literature (11) has a large increase, but its segmentation accuracy always remains low and the segmentation accuracy of literature (7) is higher compared to other algorithms in literature, but the highest does not exceed 95%, whereas the proposed algorithm always maintains the highest accuracy and is 95% when performing segmentation. Therefore, it is clear that the proposed algorithm can effectively guarantee the accuracy of image segmentation. Because the proposed algorithm uses deep reinforcement learning method, proposed algorithm used the framework of

deep learning and reinforcement learning thinking, which can segment medical images accurately.

CONCLUSION

This article proposed the 3D multimodal medical image segmentation algorithm based on deep reinforcement learning, so as to realize the segmentation of 3D multimodal medical image through deep reinforcement learning training and verify the actual performance of the algorithm by experiments. Experiments show that the algorithm in this article has high relative overlap, structural similarity, and quality factors and has strong image segmentation ability and high definition. It can effectively reduce image noise and has high application effect. In the future studies, we can continue to optimize according to the existing research theory, continuously strengthen the segmentation ability of medical images, and realize the effective segmentation of various types of medical images.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

WG and XL: conception and writing. YW: investigation and methodology. YC: data and validation. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Natural Science Foundation of Heilongjiang Province of China under grant number LH2021F039.

REFERENCES

1. Li K, Bo J, Wang P, Liu H, Shi, Y. Radar active antagonism through deep reinforcement learning: a way to address the challenge of mainlobe jamming. *Signal Processing.* (2021) 186:108130. doi: 10.1016/j.sigpro.2021.108130
2. Prianto E, Kim MS, Park JH, Bae JH, Kim JS. Path planning for multi-arm manipulators using deep reinforcement learning: soft actor–critic with hindsight experience replay. *Sensors.* (2020) 20:5911. doi: 10.3390/s20205911
3. Rabault J, Ren F, Zhang W, Tang H, Xu, H. Deep reinforcement learning in fluid mechanics: a promising method for both active flow control and shape optimization. *J Hydrodynam.* (2020) 32:234–46. doi: 10.1007/s42241-020-0028-y
4. Polvara R, Sharma S, Wan J, Manning A, Sutton, R. Autonomous vehicular landings on the deck of an unmanned surface vehicle using deep reinforcement learning. *Robotica.* (2019) 37:1867–82. doi: 10.1017/S0263574719000316
5. Liu Q, Liu Z, Xiong B, Xu W, Liu Y. Deep reinforcement learning-based safe interaction for industrial human-robot collaboration using intrinsic reward function. *Adv Eng Inform.* (2021) 49:101360. doi: 10.1016/j.aei.2021.101360
6. Lichtenegger A, Gesperger J, Kiesel B, Muck M, Eugui P, Harper DJ. Revealing brain pathologies with multimodal visible light optical coherence microscopy and fluorescence imaging. *J Biomed Opt.* (2019) 24:1–11. doi: 10.1117/1.JBO.24.6.066010
7. Roy S, Maji, P. Medical image segmentation by partitioning spatially constrained fuzzy approximation spaces. *IEEE Trans Fuzzy Syst.* (2020) 28:965–77. doi: 10.1109/TFUZZ.2020.2965896
8. Lu Z, Chen D, Xue D. Weakly supervised retinal vessel segmentation algorithm without groundtruth. *Electron Lett.* (2020) 56:1235–7. doi: 10.1049/el.2020.1893
9. Jiang J, Hu YC, Liu CJ, Halpenny D, Hellmann M, Deasy JO. Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images. *IEEE Trans Med Imaging.* (2019) 38:134–44. doi: 10.1109/TMI.2018.2857800
10. Hakim C, Achterberg M, Vernooij W. Transfer learning for image segmentation by combining image weighting and kernel learning. *IEEE Trans Med Imaging.* (2019) 38:213–24. doi: 10.1109/TMI.2018.2859478
11. Wang X, Li Z, Kang H, Huang Y, Gai, D. Medical image segmentation using pcnn based on multi-feature grey wolf optimizer bionic algorithm. *J Bionic Eng.* (2021) 18:711–20. doi: 10.1007/s42235-021-0049-4
12. Eelbode T, Bertels J, Berman M, Vandermeulen D, Blaschko MB. Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. *IEEE Trans Med Imaging.* (2020) 39:3679–90. doi: 10.1109/TMI.2020.3002417
13. Girum KB, Créhange, G, Lalande A. Learning with context feedback loop for robust medical image segmentation. *IEEE Trans Med Imaging.* (2021) 40:1542–54. doi: 10.1109/TMI.2021.3060497
14. Zhou L, Schaefferkoetter JD, Tham I, Huang G, Yan J. Supervised learning with cyclegan for low-dose fdg pet image denoising. *Med Image.* (2020) 65:101770. doi: 10.1016/j.media.2020.101770
15. Rodrigues C, Peixoto Z, Ferreira, F. Ultrasound image denoising using wavelet thresholding methods in association with the bilateral filter. *IEEE Lat Am Trans.* (2019) 17:1800–7. doi: 10.1109/TLA.2019.8986417
16. Liu X, Xu K, Zhou P, Chi J. Edge detection of retinal OCT image based on complex shearlet transform. *IET Image Processing.* (2019) 13:1686–93. doi: 10.1049/iet-ipr.2018.6634
17. Jian LA, Hang Y, Shen L, Zou Y. Ultrasound image restoration based on a learned dictionary and a higher-order MRF. *Comput Mat Appl.* (2019) 77:991–1009. doi: 10.1016/j.camwa.2018.10.031
18. Jaouen V, Bert J, Boussion N, Fayad H, Hatt M, Visvikis, D. Image enhancement with PDES and nonconservative advection flow fields. *IEEE Trans Image Process.* (2019) 28:3075–88. doi: 10.1109/TIP.2018.2881838
19. Ko Y, Moon S, Baek J, Shim H. Rigid and non-rigid motion artifact reduction in x-ray CT using attention module. *Med Image Anal.* (2021) 67:101883. doi: 10.1016/j.media.2020.101883
20. Mabrouk E, Ayman A, Raslan Y, He DR. Immune system programming for medical image segmentation. *J Comput Sci.* (2019) 31:111–25. doi: 10.1016/j.jocs.2019.01.002
21. Li X, Jiao H, Wang Y. Edge detection algorithm of cancer image based on deep learning. *Bioengineered.* (2020) 11:693–7707. doi: 10.1080/21655979.2020.1778913

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gao, Li, Wang and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fpubh.2022.862497

Decision Support System for Predicting Survivability of Hepatitis Patients

Fahad R. Albogamy[1], Junaid Asghar[2], Fazli Subhan[3,4], Muhammad Zubair Asghar[5,6], Mabrook S. Al-Rakhami[7], Aurangzeb Khan[3,8], Haidawati Mohamad Nasir[5], Mohd Khairil Rahmat[5], Muhammad Mansoor Alam[5,9,10,11,12], Adidah Lajis[5] and Mazliham Mohd Su'ud[3]*

[1] Computer Sciences Program, Turabah University College, Taif University, Taif, Saudi Arabia, [2] Faculty of Pharmacy, Gomal University, Dera Ismail Khan, Pakistan, [3] Faculty of Engineering and Computer Sciences, National University of Modern Languages-NUML, Islamabad, Pakistan, [4] Faculty of Computer and Information, Multimedia University, Kuala Lumpur, Malaysia, [5] Center for Research and Innovation, CoRI, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia, [6] Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan, Pakistan, [7] Division of Pervasive and Mobile Computing, Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, [8] Department of Computer Science, University of Science and Technology, Bannu, Pakistan, [9] Faculty of Computing, Riphah International University, Islamabad, Pakistan, [10] Malaysian Institute of Information Technology, University of Kuala Lumpur, Kuala Lumpur, Malaysia, [11] Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia, [12] Faculty of Engineering and Information Technology, School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Praveen Kumar, VIT University, India
Sheikh Tahir Bakhsh, Cardiff Metropolitan University, United Kingdom
Md. Milon Islam, Khulna University of Engineering and Technology, Bangladesh
Kadiyala Ramana, Annamacharya Institute of Technology and Sciences, India

***Correspondence:**
Mabrook S. Al-Rakhami
malrakhami@ksu.edu.sa

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *26 January 2022*
Accepted: *14 March 2022*
Published: *14 April 2022*

Citation:
Albogamy FR, Asghar J, Subhan F, Asghar MZ, Al-Rakhami MS, Khan A, Nasir HM, Rahmat MK, Alam MM, Lajis A and Su'ud MM (2022) Decision Support System for Predicting Survivability of Hepatitis Patients. Front. Public Health 10:862497. doi: 10.3389/fpubh.2022.862497

Background and Objective: Viral hepatitis is a major public health concern on a global scale. It predominantly affects the world's least developed countries. The most endemic regions are resource constrained, with a low human development index. Chronic hepatitis can lead to cirrhosis, liver failure, cancer and eventually death. Early diagnosis and treatment of hepatitis infection can help to reduce disease burden and transmission to those at risk of infection or reinfection. Screening is critical for meeting the WHO's 2030 targets. Consequently, automated systems for the reliable prediction of hepatitis illness. When applied to the prediction of hepatitis using imbalanced datasets from testing, machine learning (ML) classifiers and known methodologies for encoding categorical data have demonstrated a wide range of unexpected results. Early research also made use of an artificial neural network to identify features without first gaining a thorough understanding of the sequence data.

Methods: To help in accurate binary classification of diagnosis (survivability or mortality) in patients with severe hepatitis, this paper suggests a deep learning-based decision support system (DSS) that makes use of bidirectional long/short-term memory (BiLSTM). Balanced data was utilized to predict hepatitis using the BiLSTM model.

Results: In contrast to previous investigations, the trial results of this suggested model were encouraging: 95.08% accuracy, 94% precision, 93% recall, and a 93% F1-score.

Conclusions: In the field of hepatitis detection, the use of a BiLSTM model for classification is better than current methods by a significant margin in terms of improved accuracy.

Keywords: disease diagnosis, deep learning, hepatitis diagnostics, decision support system, bidirectional LSTM

INTRODUCTION

Data mining is becoming more common in a variety of industries, including business, education, and healthcare, as a result of the rise of AI. Many researchers are interested in healthcare decision support systems because they make it possible to glean valuable information from vast amounts of medical records. Decision support systems may be able to speed up the diagnosis process for human health care providers (1). Data mining techniques known as "deep learning" (DL) employ a variety of feature-encoding algorithms to learn from past data and deliver accurate predictions (2). Sentiment categorization (3), smart agriculture (4), and other applications are among them. In the last several years, neural network models have shown a stunning capacity to predict and classify information. Deep learning algorithms, as demonstrated by Shickel et al. (5) and Miotto et al. (6), have appeared notably in healthcare for knowledge discovery and illness diagnosis, such as heart diseases, mental disorders, and hepatitis disorders, utilizing health data.

A Need for Hepatitis Disease Prediction

Viral hepatitis is a major public health concern around the world. Hepatitis A and E viruses (HAV and HEV) are endemic in many low-income countries. They usually cause self-limiting hepatitis, but in case of compromised liver status, they can cause fulminant liver failure and chronic infection. Hepatitis B and C viruses (HBV and HCV) both cause acute liver injury and are more frequently associated with progressive liver fibrosis, cirrhosis, and a higher likelihood of liver cancer (hepatocellular carcinoma or HCC). Hepatitis D virus (HDV) is often transmitted by co-infection with HBV. Because of the severity of complications, co-infection is considered the most serious type of viral hepatitis. Hepatitis viruses are 100 times more infectious than HIV/AIDS and are the leading cause of liver cancer, which is the world's second-leading cause of cancer-related deaths. Abnormalities in liver function tests such as aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, gamma-glutamyl transferase, lactate dehydrogenase, bilirubin, and albumin can indicate the liver injury and are the integral part of the biochemical investigations in hepatitis patients. Hepatitis is one of the leading causes of liver cancer, which underscores the critical importance of early detection and diagnosis for an effective antiviral treatment (7–10).

Aims of the Project

Computational techniques like machine learning (ML) have been researched by a majority of researchers (8, 9) in recent years in order to predict the prevalence of hepatitis. Detecting hepatitis before symptoms manifested was the primary focus of their scientific work, and they were successful in this aim. Conventional encoders couldn't handle the illness dataset's predictors' relationships well. For the sake of this investigation, the suggested BiLSTM-based framework properly identifies (survivability or death) in patients with severe hepatitis. The goal of this research is to develop a prediction model that will assist physicians in determining the prognosis and survival of hepatitis patients based on medical conditions provided by the patients themselves. As a result, patients who receive early care are less likely to die.

Baselines

Akbar et al. (9) created an ML-based hepatitis disease prediction method. ML was used to better predict hepatitis using random forest (RF), support vector machine (SVM), and other classifiers. There are intrinsic links between the predictors, but this is not taken into consideration by ML classifiers. Because of this, current ML models fail to accurately forecast hepatitis patients' long-term survival using medical information. BiLSTM is an updated DL model that has previously been applied effectively in several areas, such as DDoS attack prediction (10), behavior identification (2), and others (3). To better understand hepatitis, we created a BiLSTM model. The steps are as follows: an oversampling method and a BiLSTM model were used to balance the dataset for hepatitis. Using our recommended technique, we can anticipate hepatitis patient survival from patient health data by using both data balance and the BiLSTM layer.

Problem Statement

It is challenging to reliably predict hepatitis from patient data when standard feature sets are used in conjunction with a machine learning classifier (8, 9). Aside from that, the absence of important information makes DL models for the prediction of hepatitis sickness less successful than they could be. To deal with the aforementioned challenges, the prediction of hepatitis from patient data is considered a binary-label prediction problem. The presence of hepatitis is predicted based on the disease dataset provided. The dataset contains two types of class labels: D1 (Hepatitis_patient_ survivability = yes) and D2 (Hepatitis_patient_ survivability = no). In order to predict whether or not someone has hepatitis, the neural network makes use of these two sets of information: we intend to use a deep neural network to construct an automated system that can learn from training data and predict the survivability of hepatitis patient based on contextual information.

Research Questions

The following are the research questions we hope to achieve in our study so that we can accurately diagnose hepatitis.

RO1: Using patient sickness data, the BiLSTM deep learning model will be used to create predictions regarding hepatitis infection.
RO2: BiLSTM model's comparison with traditional machine learning and deep learning models for hepatitis prediction.
RO3: The efficiency of the suggested technique in predicting hepatitis patients is compared to previous studies.

Research Contributions

The following are some of the most significant outcomes of this research work:

1. The development of a deep learning (BiLSTM) system for hepatitis disease diagnosis.
2. The proposed deep learning hepatitis detection approach outperforms existing machine learning techniques.

3. The effectiveness of the model in predicting hepatitis has been significantly improved as a result of the recommended strategy.

The following is the sequence in which the remaining sections of the study are organized: Section Related Work provides a review of the existing literature, and section Proposed Methodology outlines the approach that has been proposed by the authors. section Experimental Results and Discussion presents the findings and discussion, and Section Conclusion and Future Work discusses the application of the suggested strategy in the future.

RELATED WORK

This section summarizes previous research on hepatitis illness prediction.

Chicco and Jurman (7) used a RF classifier model to examine the EHRs of 540 control subjects and approximately 80 patients who had been diagnosed with hepatitis C. We use the best classifier (Random Forests) to find the most important variables for hepatitis C. The work of Kashif et al. (8) focuses on predicting LOLA therapy responses in hepatitis c patients. They employed K Nearest Neighbor, kStar, Bayesian Network, Randomized Forest, Radial Basis, PART, Logistic Regression, OneR, Svms, and Multi-Layer Perceptron to estimate therapeutic efficacy. For the purpose of evaluating the effectiveness of machine learning algorithms, several performance metrics are employed. It was shown that CDSS may be used to predict and diagnose Hepatitis-B infection in studies by Panigrahi et al. (11). The expert system's knowledge base has 59 rules. ES-builder, a web-based Expert System Shell, is used to implement the envisioned system (ESS). Querying the suggested system has been extensively tested to verify its efficacy. The goal of the study conducted by Wicaskno and Mudiono (12) is to use artificial intelligence to help diagnose hepatitis effectively. In this study, the certainty factor was used to diagnose the kind of hepatitis early on. Based on the data provided by the patients, this study's findings will aid in the early diagnosis of hepatitis types with precision. In order to forecast HBV integration sites, Wu et al. (13) created the DeepHBV model. DeepHBV was taught and evaluated using HBV inclusion data files from the dsVIS database by autonomously learning localized genomics features. By combining machine learning and artificial neural networks, Butt et al. (14) created an Intelligent Hepatitis C Stage Diagnosis System (IHSDS) (ANN). Only 19 of 29 attributes from the Uci repository were chosen for the study; 70% of this dataset was used for training and 30% for validation. Orooji and Kermani (15) employed machine learning to handle unbalanced data in hepatitis diagnostics. They used data preparation, data analysis, examination, and software implementation as phases. Data balance strategies were used in data preprocessing. When it comes to predicting the diagnosis of individuals with persistent hepatitis, Parisi and RaviChandran (16) have developed an innovative variable selection (FS) technique that merges neighborhood component analysis and ReliefF by taking the average of their findings, paired with a Lagrangian SVM Classifier (LSVM), which is utilized as an ML-based classification model for decision-making. Wu et al. (17) published the results of a comprehensive review and meta-analysis of the available literature, followed by validation and attention in an external cross-section of 986 cases. The area under the curve method (AUROC) and the calibrating indices were used to evaluate the model's potential to anticipate HCC within three, five, seven, and ten years. **Table 1** presents a review of selected studies.

TABLE 1 | A review of selected studies.

Study	Technique(s)	Results	Limitations
Chicco and Jurman (7)	RF classifier model	The [−1, +1] interval of MCC has increased by 8.25%	The validation cohort dataset lacked some of the attributes of the discovery cohort dataset.
Kashif et al. (8)	K Nearest Neighbor, kStar, Bayesian Network, Randomized Forest, Radial Basis, PART, Logistic Regression, OneR, Svms, and Multi-Layer Perceptron	Acc: 87%	Lack of data balancing techniques
Panigrahi et al. (11)	Web-based Expert System Shell	Knowledge base consists of 59 rules to design the expert system	Procedural knowledge can be enhanced for more effective diagnosis
Wicaksno and Mudiono (12)	certainty factor was used for early diagnosis of hepatitis	CF = 97%	Limited rule base
Wu et al. (13)	DeepHBV model	AUROC = 0.6363 AUPR = 0.5471	Lack of appropriate hidden layer selection
Butt et al. (14)	Intelligent Hepatitis C Stage Diagnosis System	Precision (94%)	Lack of external validation
Orooji and Kermani (15)	machine learning to handle unbalanced data in hepatitis diagnostics	More than 90%	Skewed dataset
Parisi and RaviChandran (16)	Merges neighborhood component analysis and ReliefF Lagrangian SVM Classifier (LSVM)	F1-score = 94%	Expanding its applicability to additional hematological diseases in order to improve patient outcomes more comprehensively.

PROPOSED METHODOLOGY

For this complex problem of decision-making, it is critical to use deep learning techniques to combine existing data and experience into a DSS. Our DSS (see **Figure 1**) incorporates data, experience,

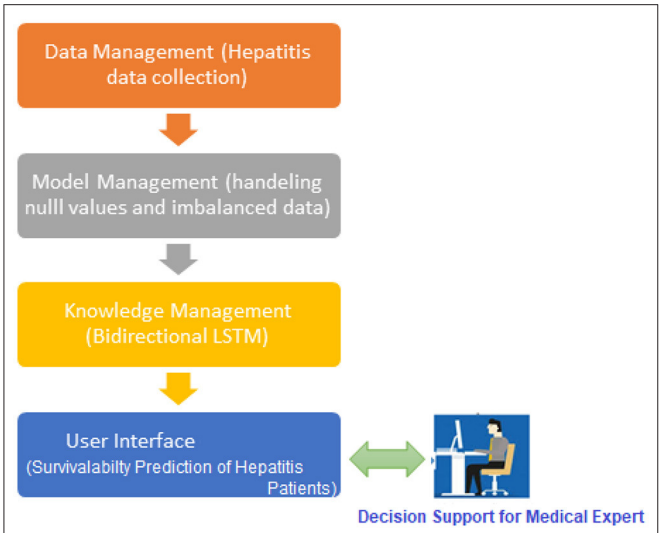


FIGURE 1 | A preview of the proposed hepatitis disease prediction system.

and models so that hepatitis clinicians may use this knowledge to form diagnostic decisions. As recommended by Turban et al., we spoke with medical experts at every phase of the design process, in order to include the users even more deeply. We followed Turban et al. (1) in building the DSS because of the DSS's nature and the sophistication of the decision topic. There are four major components of the DSS: data management, model management, knowledge-based management, and user interface.

Data Management System

A decision support system (DSS) makes use of databases and/or datasets in order to give relevant data to the decision support system. In addition to local, public, and customized sources (1), DSS data may be obtained from organizational and official sources (1). The dataset utilized in this study is the Hepatitis dataset from the University of California, Irvine repository (18). If we look at the dataset (see **Figure 2**), there are 155 instances with 20 attributes. The dataset contains 19 attributes and a single class (outcome), which may be divided into the following five groups (19).

How to Make Use of the Data

The spreadsheets are transformed into CSV format. The "csv" command line option is specified via the "pd.read" command—line interface option. This is an important Panda utility. We divided the training/testing sets through using sklearn train (80%) and test (20%) partitioning tool (2) to get the desired results.

Train Set

Approximately 80% of the data from the training set was used throughout the training process (3). Both outcome descriptors (response variables) and input parameters are included in the training phase (predictor variables).

Validation Set

Overfitting and underfitting may be handled by using validation data in the system. The model is evaluated using data from a 10% validation subset (2). When working with Keras, you have the option of changing parameters manually or automatically. Automated validation in this study will allow for a more objective evaluation of the suggested strategy.

Test Set

The test set offers unique cases to assess the algorithm's effectiveness. This process is used after considerable training and testing. Evaluation of the model is possible through the use of a test set (10). 10% of the testing data was employed, with no relation to the training instances. It is utilized once the model has already been adequately trained. It is then validated against actual data. Scikit-train-test divides the data into 90/10 ratios, with 10% of the data used for testing and the other 90% used for learning. Using a validation set, we adjusted the model's settings and analyzed the outcomes.

Treatment of Data

A 10-fold cross-validation test is used to verify the model. Ten copies of the training instance are collected and stored at each level. In this instance, we looked at one more "holdout" model. The variant with the greatest F1 score was selected for the holdout sample.

Model Management System

It performs the function of a data management system. Incorporated into DSS is a modelbase of statistical and other algorithms that allow for the provision of complex analytics. An MMS transforms data from a DMS into information by applying models to it. For the purpose of developing a viable prediction model, it is necessary to adequately pretreat the available health data. Null value substitution and imbalanced datasets are both taken care of by the data management system.

Management of Unbalanced Data Sets

There is a considerable imbalance in the underlying dataset, which treats both groups unequally. Patients who die make up 26% of the class, whereas 74% of those who live make up the remaining 26% of the class. It might be difficult to get accurate predictions from predictive algorithms when the data is imbalanced.

When a model learns from data that is skewed and unequally classified, the output generally favors the main class while neglecting the smaller categories during the classification step (20). This is seen as a class imbalance issue. Oversampling, a data processing sampling technique, is used to equalize all instances of a class. Small classes can be expanded by oversampling. Increasing the imbalance percentage is all that random up sampling accomplishes. The categorization results were much improved with the addition of small group replication. In both the T1:415 and T2:414 classes, random oversampling is applied to the balanced dataset in the same way. The total number of instances is 829.

Category	Attribute	Type	Description
Patient Profile	Age	Continuous Min=17, mean=41.2, meadian=39, max=78, Sd=12.57	Patient age
	Sex	Categorical	Male=139, female=16
Sign and Symptom	Fatigue	Categorical	No=100, yes=54, missing=1
	Malaise	Categorical	No=61, yes=93, missing=1
	Anorexia	Categorical	No=32, yes=122, missing=1
Pshysical Examination	Liver Big	Categorical	No=25, yes=120, missing=10
	Liver Firm	Categorical	No=60, yes=84, missing=11
	Spleen Palpable	Categorical	No=30, yes=120,missing=5
	Spiders	Categorical	No=,51 yes=99, missing=5
:	Ascites	Categorical	No=20, yes=130,missing=5
	Varices	Categorical	No=,18 yes=132, missing=5
Lab Test	Bilirubin	Continuous Min=0.3, mean=1.43, meadian=1.0, max=8.0, Sd=1.21	0.39, 0.80, 1.20, 2.00, 3.00, 4.00
	Alk Phosphate	Continuous Min=26, mean=105.32, meadian=85, max=295, Sd=51.51	33, 80, 120, 160, 200, 250
	Sgot	Continuous Min=14, mean=85.89, meadian=58 max=648, Sd=89.65	13, 100, 200, 300, 400, 500,
	Albumin	Continuous Min=2, mean=3.82, meadian=4 max=6.4, Sd=0.65	2.1, 3.0, 3.8, 4.5, 5.0, 6.0
	Protime	Continuous Min=0, mean=61.85, meadian=61 max=100, Sd=22.85	10, 20, 30, 40, 50, 60, 70, 80, 90
	Histology	Categorical	No=85, Yes=70
Medication Management	Steroid	Categorical	No=76, Yes=78, missing=1
	Antivirals	Categorical	No=24, Yes=131
Target Class	Die, Live	outcome	Die=32, Live=123

FIGURE 2 | Hepatitis dataset breakup.

How to Deal With Missing Data

Reliable data should be entered into the model to patch in any empty attributes (20). Filling the spaces can be done by averaging the values of the items in the previous levels, picking a random element, or switching back and using the previous level's number. Following our decision to go with the third alternative, we proceeded to update all of the missing data. In the dataset, there are just a few missing values. The PROTIME

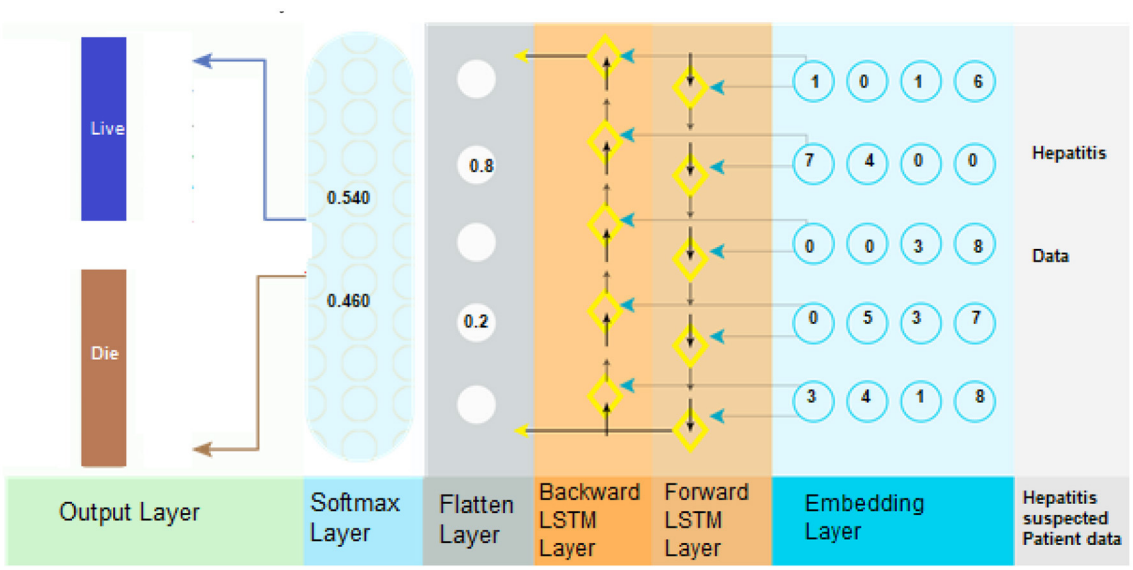


FIGURE 3 | BiLSTM is used for the hepatitis prediction system.

attribute has 67 missing values (43%), accounting for <10% of the total dataset.

The Knowledge Management System

Combined with other DSS modules, this component may be able to offer the most current information to help fix the issue. It is necessary to change information into knowledge when it has been discovered, acquired, and organized in some way. Analyzing and categorizing data is an essential part of the research process. To summarize, the suggested model consists of three main parts: embedding layer data structure, storing forward as well as backward contextual information, and softmax layer-based categorization. The second module can encapsulate features using numeric representation (21). Using Bi-LSTM to encode data contexts inside a sequence when it comes time for the final module, softmax activation is used to classify the data (see **Figure 3**). Here's a breakdown of each component:

Embedding Layer

Using embedding, it is possible to convert items (categories) into a continuous numeric array of a certain size (vectors). In neural networks, elements represent discrete variables as low-dimensional vector embeddings, which are then represented as a function of the input data. There are a number of advantages to utilizing neural embeddings in categorical data reduction and in identifying categories in the resultant space over other methods of data reduction and description. Keras facilitates the usage of embeddings by making them more accessible. Embedding vectors with attribute-level representations were employed to represent the data in the hepatitis dataset, with an embedding dimension that was equal to or greater than the dimension of the input data (22). The data embedding vector was constructed using a Keras embedding layer with 32 layers. Within the embedding layer, a two-dimensional embedding matrix (feature matrix) was built in which "represents the length of the input data" and "denotes the dimension of a word embedding" were used as input parameters. Once the embedding matrix had been produced, it was moved to the next tier in the hierarchy.

The Bi-Directional LSTM Layer

When predicting hepatitis sickness, the suggested technique makes use of a deep neural network, especially a bidirectional long-short time memory (BiLSTM), which can distinguish between D1 (Hepatitis_patient_ survivability = yes) and D2 (Hepatitis_patient_ survivability = no). Long-term dependencies are discovered through the use of the Bi-LSTM layer. As a result, it aids in the preservation of the two preceding and subsequent contexts of encoded data (23). In place of storing information from earlier contexts, a single unidirectional LSTM simply retains the information that has already been stored. As a result, Bi-LSTM is capable of analyzing encoded patient data in considerably more depth. Bi-LSTM employs both forward and backward LSTM to learn the context of data in the past and future (2). This is accomplished by the use of the following mathematical formulas (Equations 1–12).

The formulas for Forward LSTM:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (1)$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \quad (2)$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \quad (3)$$

$$c \sim t = \tau\left(W_c x_t + U_c h_{t-1} + b_c\right) \quad (4)$$

$$c_t = f_t \bigodot c_{t-1} + i_t \bigodot c \sim_t \quad (5)$$

$$h_t = o_t \bigodot \tau\left(c_t\right) \quad (6)$$

The formulas for Backward LSTM:

$$f_t = \sigma\left(W_f x_t + U_f h_{t+1} + b_f\right) \quad (7)$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t+1} + b_i\right) \quad (8)$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t+1} + b_o\right) \quad (9)$$

$$c \sim t = \tau\left(W_c x_t + U_c h_{t+1} + b_c\right) \quad (10)$$

$$c_t = f_t \bigodot c_{t+1} + i_t \bigodot c \sim_t \quad (11)$$

$$h_t = o_t \bigodot \tau\left(c_t\right) \quad (12)$$

Prediction Based on the SoftMax Algorithm

After that, SoftMax is utilized to determine the probability of forecasting target labels (i.e., the hepatitis disease). Using the formula (Equation 13), it is possible to calculate the net value of the input.

$$d_i = \sum w_i l_i + b \quad (13)$$

The input vector is denoted by "l," whereas the weight vector is denoted by "w." Bias is represented by "b." By putting it into Equation (14), one can calculate the SoftMax.

$$softmax\ \left(d_i\right) = \frac{expd^{d_i}}{\sum_{n=1}^{m} \exp^{d_n}} \quad (14)$$

Applicable Case

Based on the available illness data, we ran a number of computations to predict the presence of hepatitis. Every level of the BiLSTM model is described in depth.

Preparing the Data

For each patient occurrence in the disease dataset, our model predicts whether the patient has hepatitis D1: hepatitis Yes or "D2: hepatitis No.". Data on disease is first obtained for the DL model by using the attribute selection module. A Keras parser was used to turn the input into a matrix of indexes, which was then sent to the embedding layer of the DL model for assessment. Using the embedding layer, each illness indication is transformed into a vector of streaming values. [0.6 0.4 0.3 0.7] is an example of a vector embedding, which contains information about disease in the form of the index [1]. It ended out like this in terms of the wrapping of the matrices: [0.66, 0.53, 0.75, 0.14], [0.62, 0.71, 0.41], [0.81, 0.42, 0.74, 67], [0.54, 0.28, 0.41].

Retrieval of Contextual Data

A corrected feature map created from the previous deep network serves as input for this layer's neural network. The most important components in BiLSTM layer computations are the potential candidate value ($c \sim_t$), output gate (o_t), forget gate (f_t), and input gate (i_t).

1st Hidden Layer

(i_t) and (h_{t-1}) are the LSTM's present and previous states, respectively. To perform the calculations, we need Equations (1–6). Finally, the "$\overrightarrow{h}$" hidden state is computed (see **Figure 4**) by the first hidden layer (forward pass LSTM).

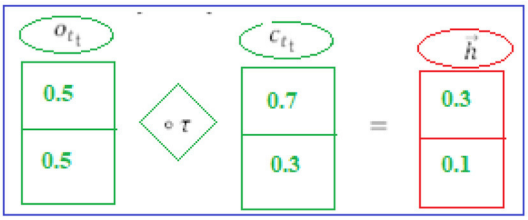


FIGURE 4 | 1st hidden layer.

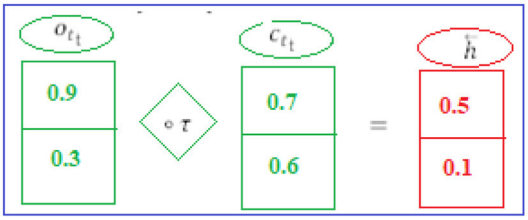


FIGURE 5 | 2nd hidden layer.

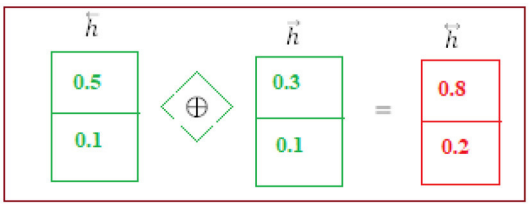


FIGURE 6 | BiSLTM outcome.

2nd Hidden Layer

The backward phase LSTM is composed (see **Figure 5**) of the current input (i_t) and the future state (h_{t+1}). The Equations (7–12) are used to do the computations. The final step is to employ the following layer to generate the hidden state $\overleftarrow{h}$ (backward phase LSTM).

BISLTM Outcome

Combining forward "$\overrightarrow{h}$" and backward "$\overleftarrow{h}$" from the neural network (see **Figure 6**), we arrive at the "$\overleftrightarrow{h}$" that make up BISLTM's final state (5).

Prediction of Hepatitis Patient Survivability

In order to find out how probable each of the tags "D1" and "D2" is, we used the SoftMax technique. The following is an example of how the overall input was calculated using Equation (13).

For Hepatitis_patient_ survivability = yes, *the class label for decision attribute 1 is "D1."*

$$d_1 = l_1 \times w_2 + l_2 \times w_2 + b$$

$$d1 = 0.6 \text{ x } 0.8 + 0.7 \text{ x } 0.2 + 0.5$$

$$d_1 = 0.48 + 0.14 + 0.5 = 1.12$$

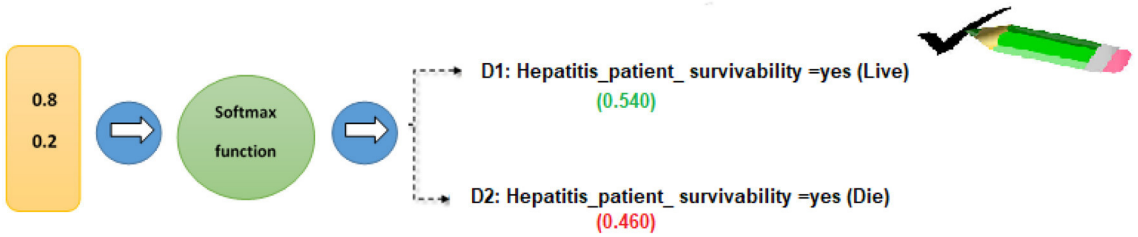


FIGURE 7 | Hepatitis patient survivability classification using the softmax function.

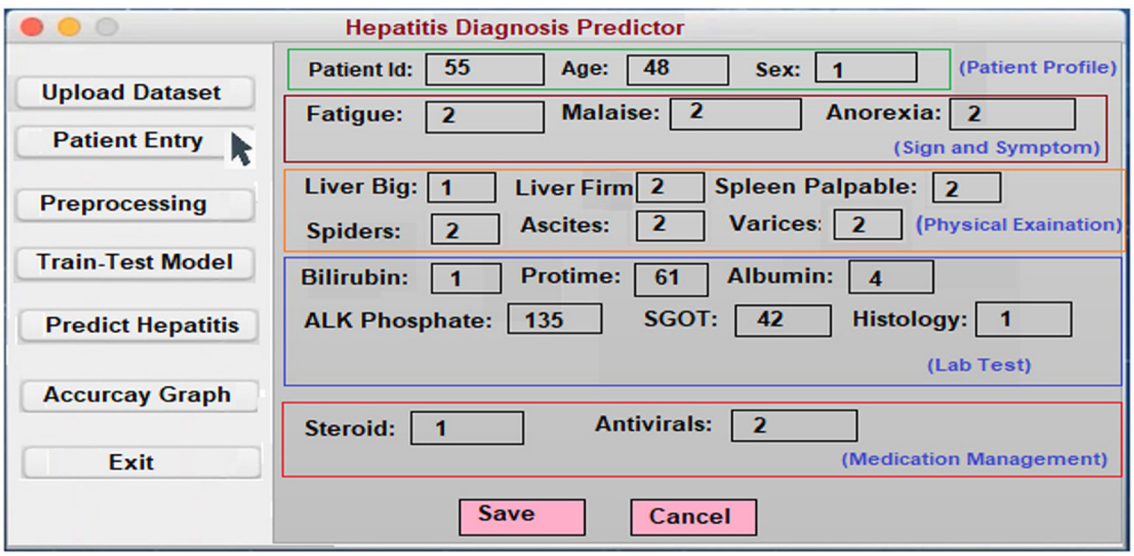


FIGURE 8 | Patient entry form.

For Hepatitis_patient_ survivability = no, *the class label for decision attribute 2 is "D2."*

$$
\begin{aligned}
d_2 &= l_1 \times w_2 + l_2 \times w_2 + b \\
d2 &= 0.5 \text{ x } 0.8 + 0.3 \text{ x } 0.2 + 0.5 \\
d2 &= 0.4 + 0.06 + 0.5 = 0.96
\end{aligned}
$$

Each target class (D1 and D2) is assigned a probability of success using the SoftMax algorithm (14).

$$
\begin{aligned}
softmax\ (d_1) &= \frac{\exp^{d_1}}{\sum_{i=1}^{n} \exp(d_1)} \\
softmax\ (d_1) &= \frac{\exp^{1.12}}{\exp^{1.12} + \exp^{0.96}} \\
softmax\ (d_1) &= \frac{3.065}{3.065 + 2.612} = \frac{3.065}{5.677} = 0.540
\end{aligned}
$$

Similarly, the SoftMax function was developed for the second class of hepatitis patient survival predictions.

$$
\begin{aligned}
softmax\ (d_2) &= \frac{\exp^{d_2}}{\sum_{i=1}^{n} \exp(d_2)} \\
softmax\ (d_2) &= \frac{\exp^{0.96}}{\exp^{1.12} + \exp^{0.96}} \\
softmax\ (d_2) &= \frac{2.612}{4.305 + 2.013} = \frac{2.612}{5.677} = 0.460
\end{aligned}
$$

Class D1: Hepatitis patient survivability = yes (live) had the highest likelihood in our calculations (0.681). Based on these data, we may estimate the patient's likelihood of survival as "D1" (live) (**Figure 7**).

User Interface

In order to assist in the prediction of hepatitis, the Keras package (3) incorporates a Python user interface. Specialists and other practitioners can use hepatitis prognosis software to better diagnose hepatitis. In order to make the content more accessible and easier to understand, it has been broken down into smaller chunks. In the software's body, there are three main parts: data acquisition, categorization, and hepatitis detection.

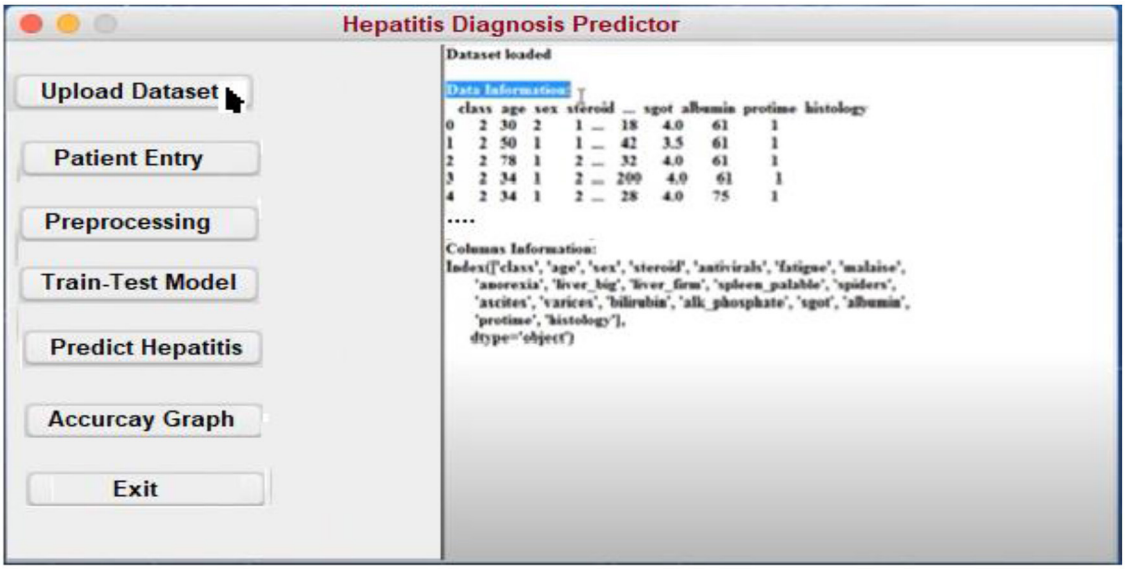


FIGURE 9 | Hepatitis prediction data input form.

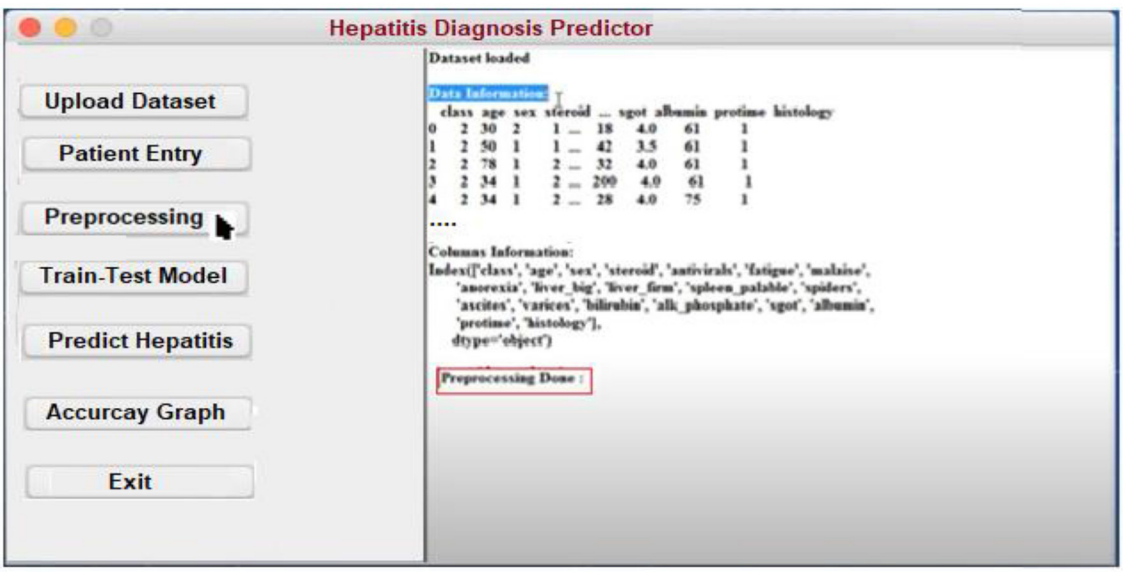


FIGURE 10 | Preprocessing form for Hepatitis prediction.

- **Patient Entry:** A patient's medical history is required for this section. Each patient in the database has a new case patient identifier created from the improved and more recent information. The hepatitis prediction system utilizes a patient's medical history to generate educated guesses about their condition. As soon as the necessary information is entered into the software, a patient specific identity is generated. Using this identifier, each patient's specific condition may be diagnosed and followed. A patient case entry is shown in **Figure 8** on the interface that has been used to gather and evaluate the data. Previous assertions are applied to the data when it is handled on the system's server. It is because of this that we have a design that can be tailored to the needs of every unique patient.
- **Uploading the Dataset:** It is possible to train classifiers and develop models using a hepatitis dataset, as shown in **Figure 9**.
- **Preprocessing:** Data preprocessing yields a classifier and a model that can predict the outcome of hepatitis. On the backend, the data is preprocessed (**Figure 10**).
- **Model Training and Testing:** The screen below shows what happens when a patient clicks on the "Model Training" option.

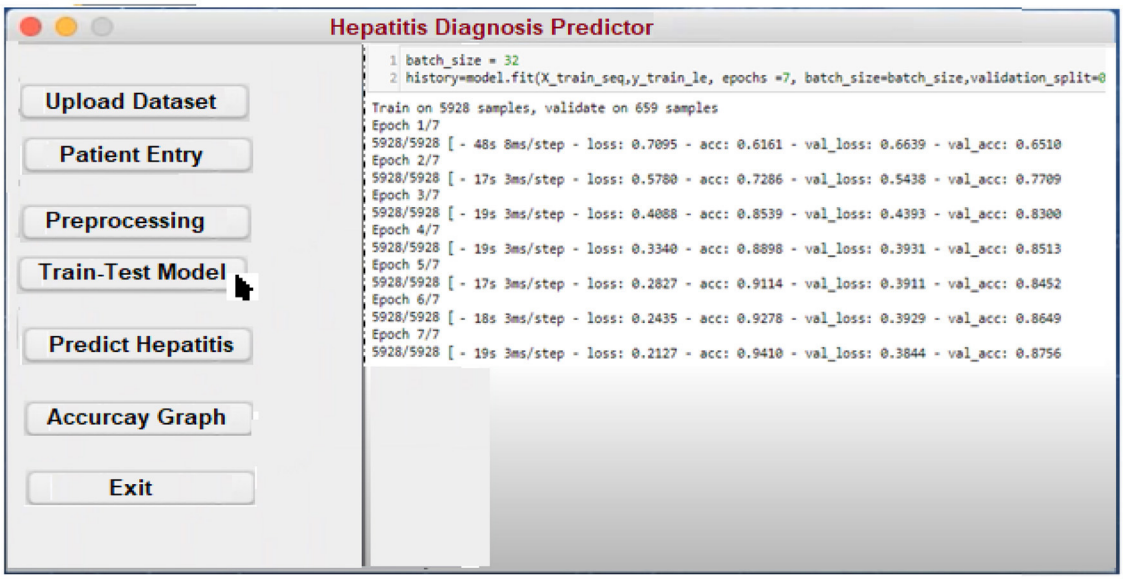


FIGURE 11 | Model training interface.

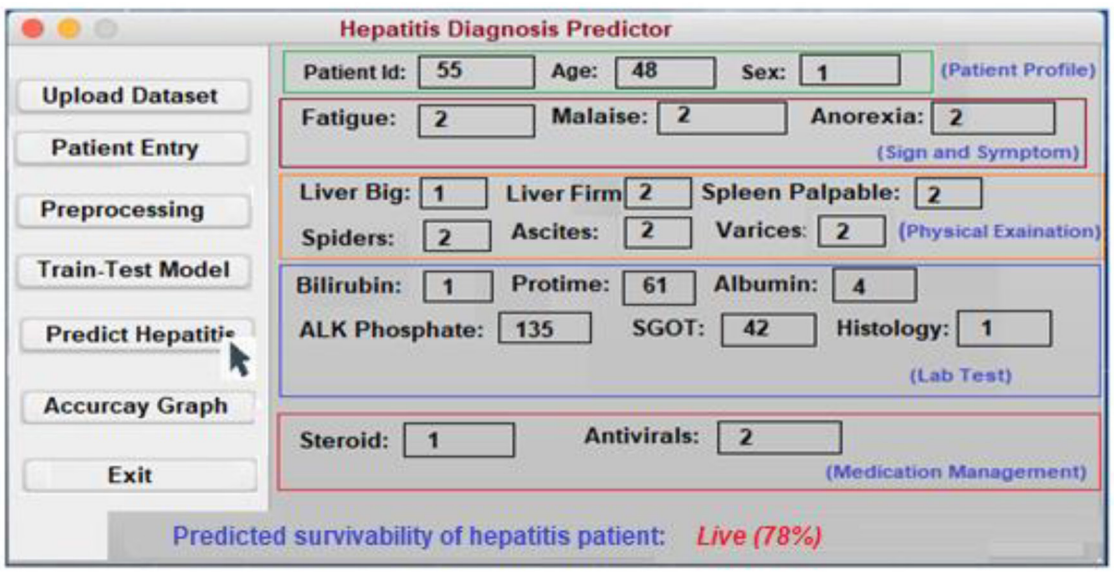


FIGURE 12 | The interface for predicting hepatitis.

It serves as a training aid by displaying the loaded data. Model training is activated by clicking on the "train-test model" (**Figure 11**).

- **Predicting survivability of Hepatitis Patient:** To predict hepatitis, all you need to do is enter the necessary patient information and click the "Predict hepatitis" option on the site. The "Update Training Set" button activates an updated training set. After inputting the patient's disease details and clicking on the "predict hepatitis" button, the findings are shown as "Hepatitis_patient_ survivability = yes," "D2: Hepatitis_patient_ survivability = no," and a predicted degree of confidence for each option. "Hepatitis_patient_ survivability = yes (Live)" is the possible outcome of the hepatitis illness prediction for the given list of criteria, as shown in **Figure 12**.

EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the results of a series of experiments aiming to address the questions raised in section Introduction.

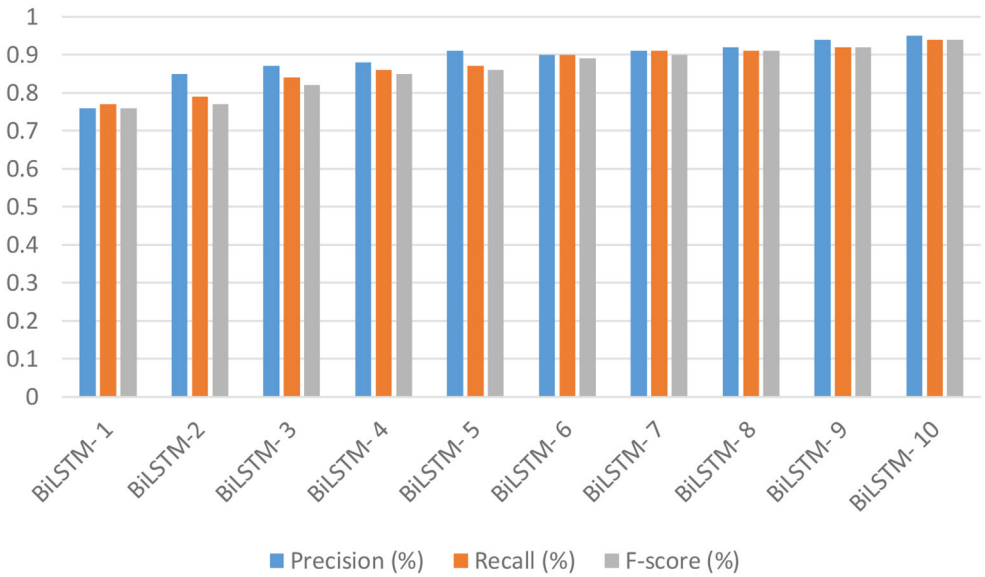


FIGURE 13 | The accuracy, recall, and f1-score of the BiLSTM deep learning models.

TABLE 2 | Accuracy, test loss, and training time of the BiLSTM models.

Model	Test accuracy	Test loss	Train time (s)
BilLSTM-1	84%	0.86	20 s
BiLSTM-2	85%	0.83	7 s
BilLSTM-3	86%	1.06	18 s
BiLSTM-4	87%	1.15	16 s
BiLSTM-5	88%	0.88	7 s
BiLSTM-6	90%	0.85	15 s
BiLSTM-7	91%	1.23	12 s
BiLSTM-8	91%	0.78	14 s
BiLSTM-9	92%	0.87	17 s
BiLSTM-10	95%	0.80	12 s

Experiment#1

Based on patient sickness data, we applied the BiLSTM deep learning model to generate hepatitis predictions.

Research question # 1 was answered by altering the parameters of one of the most commonly used models for survivability prediction of hepatitis patients, the BiLSTM algorithm. Various additional epochs and filtering techniques were employed as well. Several batch sizes and epochs are included in the algorithm's three hidden layers. As a result of using the SoftMax activation function, there were 62 vocabulary vectors, and the embedding dimension was 128 (8, 16, and 32). A selection of BiLSTM models' accuracy, recall, and F-scores are shown in **Figure 13**. This filter has an accuracy of 93% when used with the following parameters: filer size of 280, unit size 2, "f1 score" of 94%, recall of 94%, and precision of 94%. Filter number 8 has an accuracy of 95.08%.

The results of 10 BiLSTM experiments with varying parameter values are presented in **Table 2**. We compared the accuracy of each model against one another. Because it included eight filter sizes, sixteen filters, and ten LSTM units, it had the highest accuracy of all the models tested.

Experiment#2 With Datasets That Are Balanced and Unbalanced

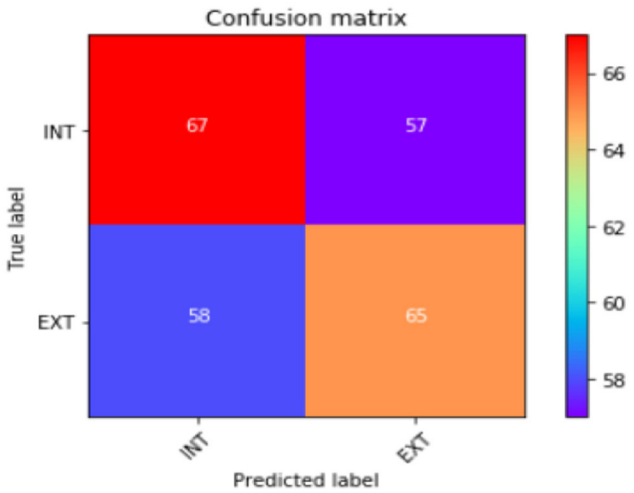
As stated in section 2.0, an imbalanced data classification, randomized training division, and testing set created an uneven class result. It is unbalanced since there are 71 surviving patients and 22 deceased. To equalize the dataset, we used random oversampling. As shown in **Figure 14**, when data balancing is used instead of not using data balancing, efficiency is significantly increased. By analyzing the empirical values, it is possible that the proposed model may be utilized to properly predict the survivability of hepatitis patients in real-world situations.

Experiment#3

This experiment is carried out to compare of the BiLSTM model for predicting hepatitis survival with traditional machine learning and deep learning. Conventional machine learning algorithms and deep neural networks were compared to see how the BiLSTM findings for diabetic illness prediction stacked up.

- **ML vs. recommended system (BiLSTM):** Patients' data was used to compare the proposed method (BiLSTM) to other commonly used machine learning methods. Machine learning makes use of feature extraction approaches like TF-IDF and CountVectorizer. **Table 3** displays the outcomes of the assessments. A brief summary of the findings is provided in the following paragraphs.
- **BiLSTM vs. SVM:** We utilized the SVM algorithm to compare the proposed BiLSTM model to various machine learning approaches. **Table 3** summarizes the findings.

```
Confusion matrix
[[67 57]
 [58 65]]
```




```
y_pred = model.predict(X_test)
y_pred = (y_pred > 0.5)
#print(y_pred)
from sklearn import metrics
print(metrics.accuracy_score(Y_test, y_pred))
```

```
0.5344129554655871
```

```
p=model.predict_proba(X_test) # to predict probability

target_names = ['INT', 'EXT']
print(classification_report(Y_test,y_pred,target_names=target_names))
```

```
              precision    recall  f1-score   support

         INT       0.54      0.54      0.54       124
         EXT       0.53      0.53      0.53       123

avg / total       0.53      0.53      0.53       247
```

FIGURE 14 | Confusion matrix.

- **BiLSTM vs. KNN:** In this experiment, we compared a typical machine learning classifier with a BiLSTM model. A closer look at **Table 3** indicates that KNN classifiers have worse precision, recall, F1-score (0.78), and accuracy (0.77) compared to other classifiers.
- **BiLSTM vs. NB:** In this experiment, the BiLSTM classifier was competing against the Nave Bayes (NB) classification. **Table 3** shows that the F1-score (0.74%), recall (0.73%), and precision (0.73%) are all lower than in the previous Table (0.73%).
- **Deep learning vs. suggested method (BiLSTM):** The proposed method is compared to existing deep learning techniques, such as long/short-term memory (LSTM), convolutional neural networks (CNN), and recurrent neural networks (RNN), in order to properly forecast hepatitis based on patient data. **Table 4** sums up the findings of the study.
- **LSTM vs. Proposed BiLSTM:** In this study, we evaluated the performance of the BILSMTM model against that of the LSTM model. **Table 4** shows that the LSTM model had the lowest precision, recall, F1-score, and accuracy of the four models tested.
- **CNN vs. Proposed BiLSTM:** Our goal was to compare the performance of the BiLSTM approach to that of the CNN model in this particular experiment. For accuracy, precision,

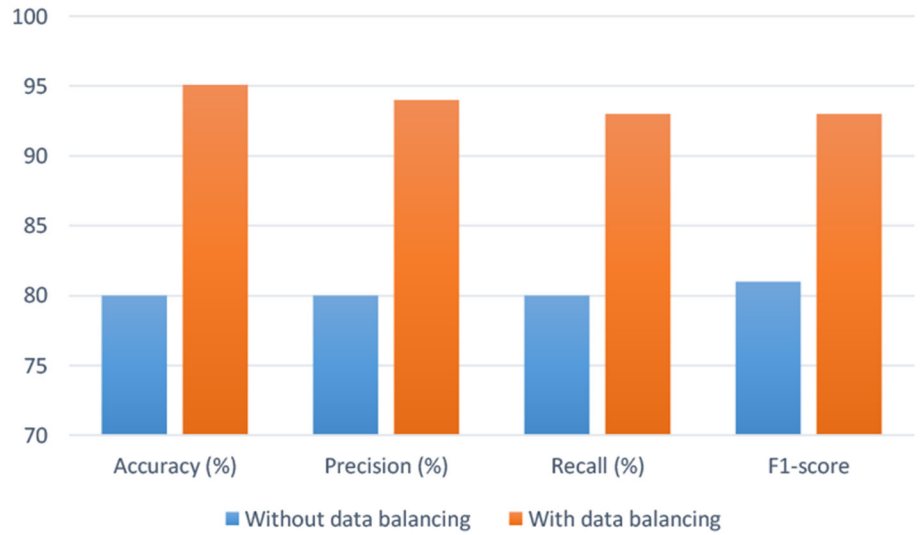


FIGURE 15 | Proposed model's performance with and without balancing data.

recall, F1, and precision, the CNN model behaved poorly in **Table 4**.

- **RNN vs. Proposed BiLSTM:** We ran this test to see which approach worked best. On the basis of the results in **Table 4**, it can be concluded that the RNN model's performance on these four metrics is less than ideal.

Experiment#4

In this experiment, the suggested method's efficacy in predicting hepatitis patients is compared to baseline research. In the third research question, we evaluated the suggested BiLSTM model's effectiveness in similar research. To determine how effective the proposed system is, it is contrasted with a variety of other comparing methodologies. According to this study (**Figure 15**), BiLSTM surpasses a baseline study when compared to the latter. For a variety of reasons, it is difficult to conduct a comprehensive assessment of published methodologies. These approaches were hard to compare because of the variety of datasets.

Kashif et al. (8)'s work: Kashif et al. (8) created an ML-based hepatitis disease prediction method. ML was used to better predict hepatitis using random forest (RF), support vector machine (SVM), and other classifiers

Akbar et al. (9)'s work: Prediction of hepatitis using ML has been proposed. Hepatitis data sets were analyzed using a variety of machine learning methodologies. The performance of a DL model can be improved by using a more effective data balancing method.

Proposed work (our model): The DL-based hepatitis prediction approach uses a deep neural network. **Figure 16** shows that the predictor variables (**Figure 16**) selected have a considerable influence on the anticipated (target) parameter. The combination of data balance and the BiLSTM neural learning model is the most important factor in our achievement in predicting hepatitis illnesses. By using the BiLSTM layer, contextual information can be retained.

TABLE 3 | Classifiers based on machine learning vs. suggested model (BiLSTM).

ML modal	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
SVM	78	79	80	79
KNN	77	78	77	78
NB	74	73	73	73
Proposed (BiLSTM)	95.08	94	93	93

TABLE 4 | BiLSTM in comparison to other deep learning models.

DL model	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
LSTM	85.14	86	85	85
CNN	85.24	83	84	83
RNN	84.15	82	82	82
Proposed (BiLSTM)	95.08	94	93	93

Analyzing Results

A comparison is made between experts' predictions and the prognosis supplied by the suggested strategy, and the performance of the proposed method is evaluated. The processes for the first 12 patients are depicted in **Table 5**.

CONCLUSION AND FUTURE WORK

Because of the tremendous expansion in healthcare content, it has become more important to gather and analyze healthcare data in order to discover hepatitis illness in patients. A successful DSS model based on DL was created and put into practice in order to accomplish this. A deep neural network (BiLSTM) is used to predict hepatitis based on data gathering, preprocessing,

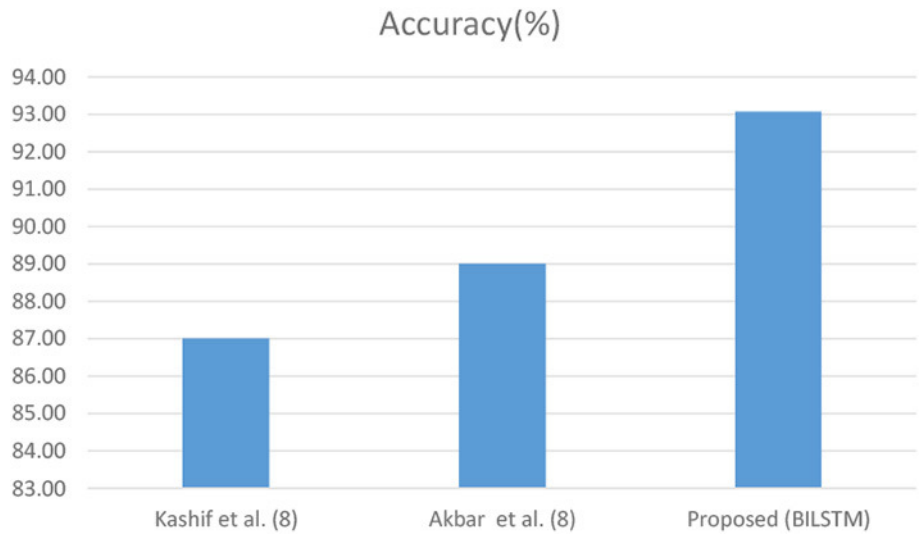


FIGURE 16 | Comparison of other research and the BiLSTM model.

TABLE 5 | Human expert vs. proposed system prognosis.

Suspected hepatitis patient	Diagnosis by hepatitis expert	Prediction by BiLSTM (proposed)
1	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
2	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
3	Hepatitis_patient_ survivability = no	Hepatitis_patient_ survivability = no
4	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = no
5	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
6	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
7	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
8	Hepatitis_patient_ survivability = no	Hepatitis_patient_ survivability = no
9	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
10	Hepatitis_patient_ survivability = no	Hepatitis_patient_ survivability = no
11	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes
12	Hepatitis_patient_ survivability = yes	Hepatitis_patient_ survivability = yes

and validation. Additional experiments were conducted using both balanced and unbalanced data sets. A BiLSTM model was also utilized to predict the chance of having hepatitis in the future. The results are encouraging when compared to prior attempts. The suggested strategy's dependence on embedding rather than on a pre-trained model is an obvious limitation. In the future, the use of pre-trained algorithms like word2vec or Fasttext with hepatitis data sets from different fields (e.g., patient data from different fields) may be looked into.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

FA and MAs: conceptualization. MAs and FS: methodology. MA-R and JA: software. FA and JA: validation. AK and HN: formal analysis. MR: investigation. MAs: resources. MAl and JA: data curation. MAs: writing—original draft preparation. JA: writing—review and editing. MAs, MS, and JA: visualization. AL: supervision and project administration. FA: funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

The APC was funded by Taif University Researchers Supporting Project Number (TURSP-2020/331), Taif University, Taif, Saudi Arabia.

ACKNOWLEDGMENTS

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs. The authors would also like to acknowledge the support from Taif University Researchers Supporting Project Number (TURSP-2020/331), Taif University, Taif, Saudi Arabia.

REFERENCES

1. Turban E, Aronson JE, Liang TP, MacCarthy RV. *Decision Support Systems and Intelligent Systems*. Upper Saddle River, NJ: Pearson Prentice-Hall (2005).
2. Ahmad H, Asghar MU, Asghar MZ, Khan A, Mosavi AH. A hybrid deep learning technique for personality trait classification from text. *IEEE Access.* (2021) 9:146214–32. doi: 10.1109/ACCESS.2021.3121791
3. Khattak A, Habib A, Asghar MZ, Subhan F, Razzak I, Habib A. Applying deep neural networks for user intention identification. *Soft Comput.* (2021) 25:2191–220. doi: 10.1007/s00500-020-05290-z
4. Rose DC, Sutherland WJ, Parker C, Lobley M, Winter M, Morris C, et al. Decision support tools for agriculture: towards effective design and delivery. *Agric Syst.* (2016) 149:165–74. doi: 10.1016/j.agsy.2016.09.009
5. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* (2017) 22:1589–604. doi: 10.1109/JBHI.2017.2767063
6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* (2018) 19:1236–46. doi: 10.1093/bib/bbx044
7. Chicco D, Jurman G. An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis. *IEEE Access.* (2021) 9:24485–98. doi: 10.1109/ACCESS.2021.3057196
8. Kashif AA, Bakhtawar B, Akhtar A, Akhtar S, Aziz N, Javeid MS. Treatment response prediction in hepatitis C patients using machine learning techniques. *Int J Technol Innov Manag.* (2021) 1:79–89. doi: 10.54489/ijtim.v1i2.24
9. Akbar W, Wu WP, Saleem S, Farhan M, Saleem MA, Javeed A, et al. Development of hepatitis disease detection system by exploiting sparsity in linear support vector machine to improve strength of AdaBoost ensemble model mobile information systems. *Personal Commun Technol Smart Space.* (2020) 2020:8870240. doi: 10.1155/2020/8870240
10. Alghazzawi D, Bamasaq O, Ullah H, Asghar MZ. Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection. *Appl Sci.* (2021) 11:11634. doi: 10.3390/app112411634
11. Panigrahi N, Ayus I, Jena OP. An expert system-based clinical decision support system for hepatitis-b prediction and diagnosis. *Mach Learn Healthcare Appl.* (2021) 57–75. doi: 10.1002/9781119792611.ch4
12. Wicaksno AP, Mudiono DRP. Early detection of hepatitis by using certainty factor. In: *The First International Conference on Social Science, Humanity, and Public Health (ICOSHIP 2020)*. Atlantis Press. p. 93–7
13. Wu C, Guo X, Li M, Shen J, Fu X, Xie Q, et al. DeepHBV: a deep learning model to predict hepatitis B virus (HBV) integration sites. *BMC Ecol Evol.* (2021) 21:1–10. doi: 10.1186/s12862-021-01869-8
14. Butt MB, Alfayad M, Saqib S, Khan MA, Ahmad M, Khan MA, et al. Diagnosing the stage of hepatitis C using machine learning. *J Healthc Eng.* (2021) 2021:8062410. doi: 10.1155/2021/8062410
15. Orooji A, Kermani F. Machine learning based methods for handling imbalanced data in hepatitis diagnosis. *Front Health Inform.* (2021) 10:57. doi: 10.30699/fhi.v10i1.259
16. Parisi L, RaviChandran N. Syncretic feature selection for machine learning-aided prognostics of hepatitis. *Neural Process Lett.* (2021) 2021:1–25. doi: 10.1007/s11063-021-10668-7
17. Wu S, Zeng N, Sun F, Zhou J, Wu X, Sun Y, et al. HCC prediction models in chronic hepatitis B: a systematic review of 14 models and external validation. *Clin Gastroenterol Hepatol.* (2021) 19:2499–513. doi: 10.1016/j.cgh.2021.02.040
18. UCI Machine Learning Repository: Hepatitis Data Set. (2022). Available online at: https://archive.ics.uci.edu/ml/datasets/hepatitis (accessed January 17, 2022).
19. Goh M. Predicting Hepatitis Patient Survivability (Uci dataset). Medium. Available online at: https://towardsdatascience.com/predicting-hepatitis-patient-survivability-uci-dataset-71982aa6775d (accessed November 23, 2020)
20. Khan AS, Ahmad H, Asghar MZ, Saddozai FK, Arif A, Khalid HA. Personality classification from online text using machine learning approach. *Int J Adv Comput Sci Appl.* (2020) 11:58. doi: 10.14569/IJACSA.2020.0110358
21. Alazab M, Khan S, Krishnan SSR, Pham QV, Reddy MPK, Gadekallu TR. A multidirectional LSTM model for predicting the stability of a smart grid. *IEEE Access.* (2020) 8:85454–63. doi: 10.1109/ACCESS.2020.2991067
22. Gadekallu TR, Iwendi C, Wei C, Xin Q. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021) 647–58. doi: 10.1049/ipr2.12222
23. Bhattacharya S, Maddikunta PKR, Hakak S, Khan WZ, Bashir AK, Jolfaei A, et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimed Tools Appl.* (2020) 1–25. doi: 10.1007/s11042-020-09988-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors MAs.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Albogamy, Asghar, Subhan, Asghar, Al-Rakhami, Khan, Nasir, Rahmat, Alam, Lajis and Su'ud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fpubh.2022.885212

Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection

Shivaji D. Pawar[1,2], Kamal K. Sharma[3], Suhas G. Sapate[4], Geetanjali Y. Yadav[5], Roobaea Alroobaea[6], Sabah M. Alzahrani[6] and Mustapha Hedabou[7]*

[1] Department of Computer Science and Engineering, Lovely Professional University, Jalandhar, India, [2] SIES Graduate School of Technology, Navi Mumbai, India, [3] School of Electronics and Electrical Engineering, Lovely Professional University, Jalandhar, India, [4] Department of Computer Science and Engineering, Annasaheb Dange College of Engineering and Technology, Sangli, India, [5] NMC Royal Medical Centre Karama, Abu Dhabi, United Arab Emirates, [6] Department Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, [7] School of Computer Science, Mohammed VI Polytechnic University, Ben Guerir, Morocco

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Clement Nyirenda, University of the Western Cape, South Africa
Xiao-Zhi Gao, University of Eastern Finland, Finland

****Correspondence:***
Kamal K. Sharma
kamal.23342@lpu.co.in

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *27 February 2022*
Accepted: *14 March 2022*
Published: *25 April 2022*

Citation:
Pawar SD, Sharma KK, Sapate SG, Yadav GY, Alroobaea R, Alzahrani SM and Hedabou M (2022) Multichannel DenseNet Architecture for Classification of Mammographic Breast Density for Breast Cancer Detection. Front. Public Health 10:885212. doi: 10.3389/fpubh.2022.885212

Percentage mammographic breast density (MBD) is one of the most notable biomarkers. It is assessed visually with the support of radiologists with the four qualitative Breast Imaging Reporting and Data System (BIRADS) categories. It is demanding for radiologists to differentiate between the two variably allocated BIRADS classes, namely, "BIRADS C and BIRADS D." Recently, convolution neural networks have been found superior in classification tasks due to their ability to extract local features with shared weight architecture and space invariance characteristics. The proposed study intends to examine an artificial intelligence (AI)-based MBD classifier toward developing a latent computer-assisted tool for radiologists to distinguish the BIRADS class in modern clinical progress. This article proposes a multichannel DenseNet architecture for MBD classification. The proposed architecture consists of four-channel DenseNet transfer learning architecture to extract significant features from a single patient's two a mediolateral oblique (MLO) and two craniocaudal (CC) views of digital mammograms. The performance of the proposed classifier is evaluated using 200 cases consisting of 800 digital mammograms of the different BIRADS density classes with validated density ground truth. The classifier's performance is assessed with quantitative metrics such as precision, responsiveness, specificity, and the area under the curve (AUC). The concluding preliminary outcomes reveal that this intended multichannel model has delivered good performance with an accuracy of 96.67% during training and 90.06% during testing and an average AUC of 0.9625. Obtained results are also validated qualitatively with the help of a radiologist expert in the field of MBD. Proposed architecture achieved state-of-the-art results with a fewer number of images and with less computation power.

Keywords: breast cancer, BIRADS Density Classification, DenseNet, deep learning, multichannel architecture, mammographic breast density

INTRODUCTION

Breast cancer has its reputation as a deadly disease and it is the second most frequent event of departure for the women community in society. It is frequently diagnosed in women with 12.3% compared to the average population (1–4). The incidence of breast cancer is associated with many biomarkers such as calcifications, masses, mammographic breast density (MBD), and architectural distortion. Advanced apprehension and proactive strategy are the unique alternatives to protect the lives of breast cancer cases and regress their subconscious shock (5–8). MBD is an essential biomarker in interpreting digital mammograms and preparing a systematic mammography screening program. According to epidemiological investigations, females with highly dense breast tissue risk developing breast malignancy (9, 10). MBD is a radiographically visible density on the mammogram, consisting of lobular elements, ducts, and fibrous connective tissue compared to the lucent fatty tissue in the breast (11). The reports are prepared during the digital mammography screening program as per the American College of Radiology Breast Imaging Reporting and Data System (ACR BIRADS) catalog. This catalog was last modified in November 2015 (12, 13). As per the BIRADS, MBD gets classified into the four significant groups of mammographic breast density as class A, class B, class C, and class D, which reduces the sensitivity of digital mammography (14). **Figure 1** depicts all the MBD BIRADS classes.

Different scientific studies have revealed that digital mammograms' sensitivity strongly depends on the density class of the breast tissue. In dense breasts, the sensitivity of mammograms is as low as 63%, while in the low-density breast, there is an exponential rise of 87%. Hence, patients with high-density breasts have to go for additional imaging such as tomosynthesis, ultrasound, or breast MR to enhance cancer detection chances (15). Researchers have proposed many semiautomatic and automatic approaches for measuring breast density from the last two decades. However, the assessment of mammographic breast density (MBD) is subjective, which expert radiologists do (16). Despite this labeling, due to poor interreader and intrareader reproducibility, MBD classification has many limitations (17).

For an automatic objective assessment of MBD classification, many study efforts have been in progress from the last few decades. The initial study focuses on image processing techniques such as area-based thresholding, region growing, and clustering algorithms. A significant step formulated was the emergence of the machine learning (ML) algorithms based on different extracted image topographies from the histogram, texture intensities, patterns, and image acquisition parameters. Nowadays, deep learning algorithms provide yet another leap forward in MBD classification. Deep learning algorithms can go substantially deeper and discover all the significant features from the image. Due to system architecture and hardware improvements, it is possible to train deep learning architecture intensely. This improvement makes deep learning architecture an excellent tool for medical image analysis. Many deep learning architectures such as LeNet, (visual geometry group) VGG19, highway networks, residual networks, and DenseNet are recorded in literature.

Nevertheless, before every deep learning network lets more intelligence, a new study intricacy happens, the "vanishing gradient problem." DenseNet (dense convolutional network) provides unique insight to secure the best data flow between layers to solve the connectivity problem. This interface immediately combines all the layers, agreeing on characteristic map dimensions in feed-forward type; hence, the individual layer receives input from all the previous layers also transfers its distinct map to all the farther layers. Thus, the DenseNet concatenate feature map passes through all the subsequent layers instead of summarizing features such as ResNet. This concept includes L (L+1)/2 connections instead of L connections, identifying a dense connectivity pattern (18). The consequence of this connectivity guide, i.e., DenseNet architecture, provides the following advantages:

1. Effective solution for gradient vanishing.
2. Consolidation in characteristic distribution.
3. Provision of feature recycle.
4. Significant reduction in training parameters.
5. Easy to train and offers better parameter efficiency.

Due to these advantages, it is helpful to use this model without pretraining for medical image analysis.

The elemental aspiration behind this study is to investigate the use of multichannel DenseNet architecture for MBD classification and analyze the proposed architecture compared with different existing methods. **Figure 2** depicts the proposed multichannel architecture for MBD classification.

The key feature of this architecture is classifier performance that is evaluated with 800 digital mammograms among the diverse BIRADS density categories. Results are confirmed exclusively by expert radiologists and objectively with classification accuracy and the area under the curve (AUC). Various sections described in this article are as below. Related study section covers unique existing deep learning algorithms for MBD classification, Study dataset section presents the aspects of the source dataset, Proposed methodology section represents the proposed architecture, and Analysis of experiment and outcomes section introduces the empirical outcomes. Finally Discussion section discusses the future scope and Segment 7 concludes this study.

RELATED STUDY

Mammographic breast density classification is a long-lasting study area due to more thought-provoking and challenging image preprocessing, segmentation, and classification tasks. Successful deep learning classifiers such as ResNet, VGGNet, and GoogLeNet give brand-new pathological imaging and investigation prospects. This section describes the remarkable of such existing methods used for MBD classification. In summation, these classifiers offer better results on different imaging modalities for image classification. Intrinsically, high interreader fluctuations are the prime problem in MBD

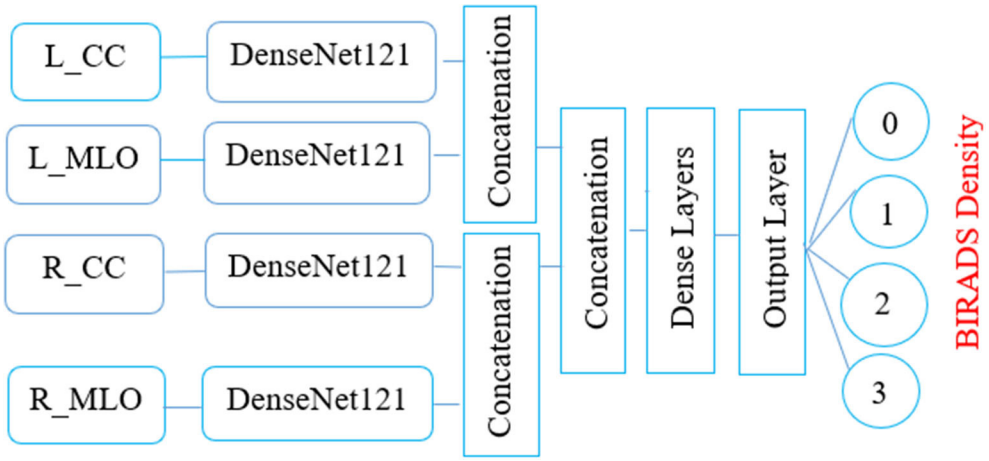


FIGURE 1 | The proposed multichannel architecture for mammographic breast density classification.

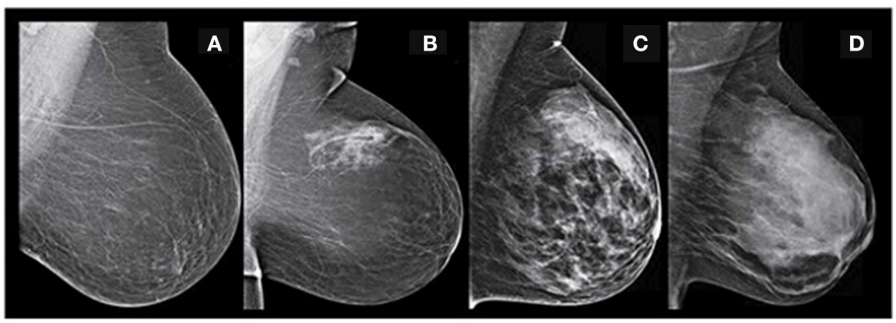


FIGURE 2 | BIRADS classification-**(A)** fatty-class A **(B)** fat with some fibro glandular tissue -class B **(C)** heterogeneous dense-class C **(D)** extremely dense-class D (Image courtesy: Densebreast-info.org).

assessment. To overwhelm this problem, Ciritsis et al. (19) suggested deep learning architecture with 11 convolutional layers, three fully connected layers, and an output SoftMax layer to distinguish breast density into two classes and four classes. Two expert radiologists marked density ground truth during this classification task. A combination of mediolateral oblique (MLO) and craniocaudal (CC) (left or right) views (20,578) digital mammograms were used to train the proposed model. During the training, to enhance the model's performance and desist overfitting, the convolutional layers are zero-padded and used a dropout rate of 50%. Batch size and the highest number of epochs used during training are 40 and 120, respectively. Input data segmented as 70% for training and 30% for validation. This model is validated independently on CC and MLO views for two-class classification and achieved an overall classification accuracy of 89.9 and 86.6%.

Predominantly, fat appears darker than fibroglandular tissue in MBD assessment. Hence, the pixel intensity of the histogram acts as an essential feature for classifier training. Wu et al. (20) suggested this technique for MBD classification. In this method, the SoftMax regression classifier is used to learn pixel intensity histograms. Four regular views [left CC (LCC), right CC (RCC), left MLO (LMLO), and right MLO (RMLO)] of 2,00,000 screening images were given individually to train the prototype. The proposed architecture is deep and consists of 100 hidden units between input and output layers. All the hidden layers use Rectified Linear Unit (ReLU) as an activation unit. This model is tested for two- and four-class classification and achieved 81.1 and 82.5% classification accuracy. Thus, this model provides moderate classification accuracy despite an extensive dataset. Lizzi et al. (21) suggested a residual convolutional network for MBD classification. The recommended design consists of 41

TABLE 1 | Input dataset used for testing and validation of proposed algorithm.

BI_RADS Density Class	Number of images
Class-A	200
Class-B	200
Class-C	200
Class-D	200
Total	800

convolutional layers formed in residual blocks with 2 million parameters. The first block of this architecture consists of a convolutional layer, a batch normalization layer, a leaky ReLU as an activation function, and a two-dimensional (2D) max pooling. A series of four sections consisting of three residual modules uses the output features of the input block. Leaky ReLU with $\alpha = 0.2$ activation functions was used to train the architecture. Categorical cross-entropy as a loss function validates the performance of the model. Maximum accuracy with four-class classification is 78% and the two-class accuracy is 89.4%.

In a comparative study for evaluating the performance of deep learning and transfer learning on a similar dataset, Lee et al. (17) proposed both the approaches on 22,000 mammographic images. In this approach, expert radiologists marked density ground truth for the input images. Initially, the model's training starts with 500 images with the AUC of 0.942 and the final value of the AUC is 0.9882 reported on the whole dataset. Then, the proposed model is cross-validated with the transfer learning (ImageNet) application and obtained the overall AUC of 0.9857. Thus, this study shows that both the deep learning and transfer learning applications provide almost identical results on the equivalent dataset. The requirement of a larger dataset for training is an essential need of deep learning architecture. Shi et al. (22) proposed optimized lightweight deep learning architecture to optimize deep learning performance on smaller datasets. This architecture combines three convolutional neural networks (CNNs), one dense layer, and an output layer with the SoftMax function. Data augmentation is done with additional image processing to increase the numbers in the dataset. This architecture was trained and tested on the 322 mini-Mammographic Image Analysis Society (MIAS) dataset. This architecture provides overall accuracy of 83.6% on four-class classification. The main limitation of lightweight architecture on smaller datasets is the low stability of the network, which may occasionally cause large and significant variations in the accuracy. Data augmentation can enlarge the dataset in this method, but it is still challenging to get well-trained convolutional layers due to little diversity between the original and generated datasets. In the literature, there are two ways that are recorded to enhance the model performance with a smaller dataset, which are generative adversarial network (GAN) and another is transfer learning (23).

Recently, different researchers proposed the concept of transfer learning on a smaller dataset. For example, Kaiser et al. (24) proposed new architecture that can take all the four views of single patients to classify MBD into two-class classification (dense and nondense). For this purpose, the author proposed four-channel VGGNet architecture to extract all the features with average global pooling from input mammograms. Before the classification layer, to concatenate all the input layer features, two dense layers are used. Then, the proposed model is trained with 5-fold cross-validation and recorded 88% classification accuracy with the AUC of 0.954. Finally, subjective assessment is done with a panel of 32 radiologists to compare interobserver variability. In this approach, interobserver variability for breast density assessment is observed even high in two-class classification. Thus, the automated processes for MBD can help to minimize interobserver variability. Despite different automated approaches, MBD assessment is subjective and consists of intra- and interobserver variations. Objective evaluation of other commercially viable methods consists of mixed evaluation results. Another fundamental limitation is that most of existing deep learning methods need a higher dataset and validated density ground truth; hence, data acquisition becomes difficult for researchers. In addition, mammographic images are vendor specific, making deep learning more robust; training the deep learning model through different vendor-specific samples is required, another bottleneck in MBD classification. All the limitations mentioned above result in moderate objective MBD classification accuracy.

The primary motivation behind this study is to investigate the transfer learning application of DenseNet architecture toward enhancing the classification accuracy of MBD. A significant contribution is the design and development of multichannel architecture to utilize four mammographic views of a single patient.

STUDY DATASET

The intended study utilizes the openly available dataset from digital database for screening mammography (DDSM) (25), consisting of 2,620 samples of various classes labeled as benign, normal, and malignant with confirmed pathogeny data. The aimed algorithm uses 200 Right-MLO, 200 Left_MLO, 200 R_CC, and 200 L_CC views. A total of 800 mammograms are used for training and testing purposes. The ground truth of each class is labeled with the help of specialist radiologists team into four classes as 0, 1, 2, and 3 as a four MBD classes. All the density groups consist of 200 cases of different images (MLO, LMO, R_CC, and L_CC). **Table 1** presents the details of the ground truth input dataset used in this proposed study.

PROPOSED METHODOLOGY

This segment explains the proposed MBD classification technique and divided into three subsections as presented subsequently.

Segmentation of Pectoral Muscle

Instead of using raw images for models' training, the proposed method preprocesses both the LMO and MLO views. Input mammographic images consist of high-intensity marks of artifacts and tags and pectoral muscle. Those fields in the breast region can reduce the MBD classification accuracy (26, 27). The proposed method uses the depth-first search algorithm, our previous study (28), to remove pectoral muscle, artifacts, and tags from all the MLO and LMO views. This preprocessing study helps the model to classify all the BIRADS classes correctly with less input images. Depth-first search (DFS) algorithm identify all the unwanted high-intensity areas (artifacts and pectoral muscle) of MLO and CC views and removes them, which further help the model to make the correct decision. **Figures 3**, **4** depict the input and output images after preprocessing.

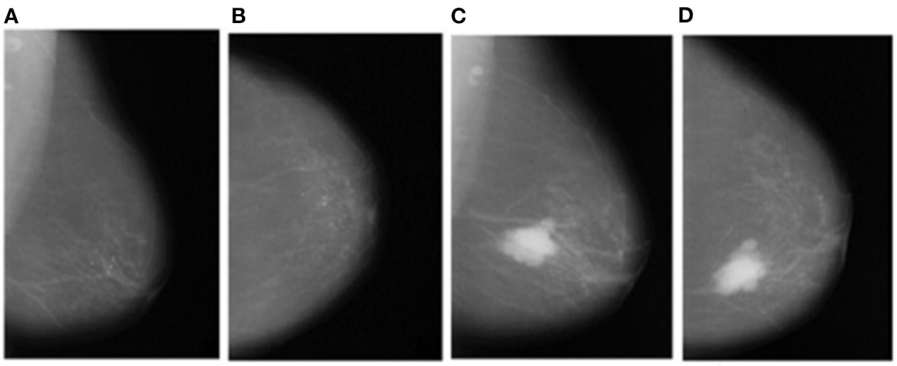


FIGURE 3 | Input raw images- **(A)** Left_MLO **(B)** Left_CC **(C)** Right_MLO **(D)** Right_CC.

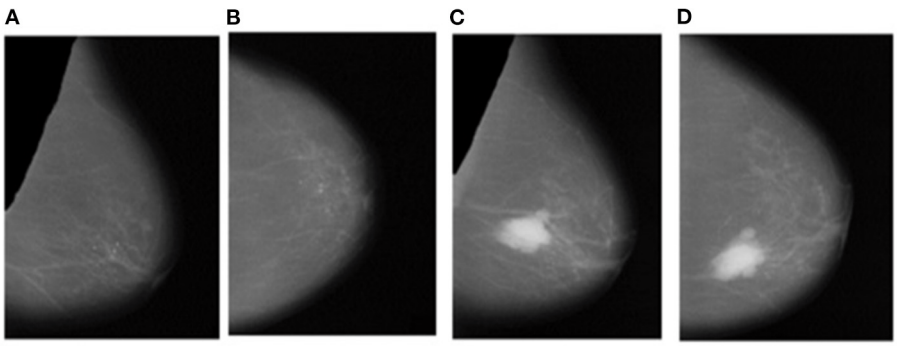


FIGURE 4 | Output Images after segmentation and cropping - **(A)** Left_MLO **(B)** Left_CC **(C)** Right_MLO **(D)** Right_CC.

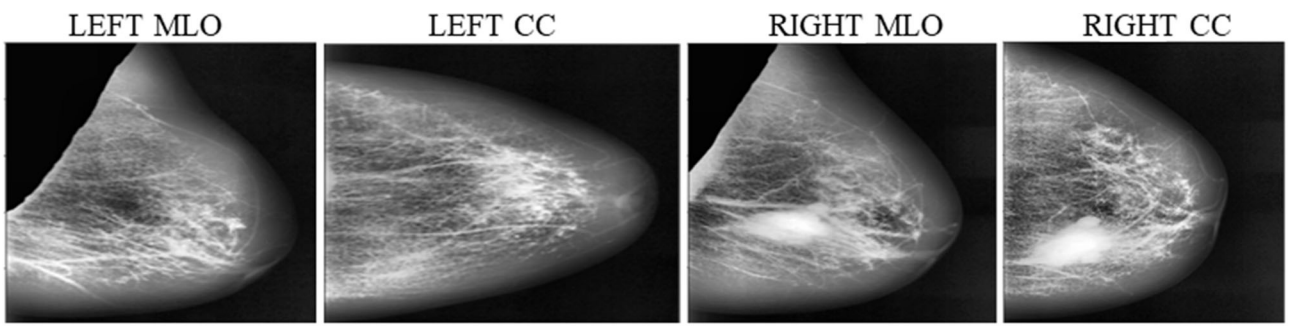


FIGURE 5 | Contrast enhancement of input images.

Contrast Enhancement

Mammographic breast density classification is a function of the density of fibroglandular tissues inside the breast. Contrast enhancement helps to improve the visibility of fibroglandular tissues. Subsequently, it helps to improve the classification accuracy of deep learning models. In literature, many contrast enhancement methods are recorded to enhance the quality of medical images. These methods are "histogram equalization (HE)," "adaptive histogram equalization (AHE)," and "wavelet transform (WT) coefficients" (29–31). But, these documented practices take ample processing time and are less effective in noise reduction. Another recorded method is the "unsharp masking (USM) method," which enhances the local contrast inside the image by limiting the global contrast. Still, this technique creates artifacts in the image. Due to this, the image looks artificial; therefore, it is not fitting for the enrichment of the medical images (32–34).

In contrast to gain agreement for the high-frequency element of the image, which is the basic principle behind "adaptive contrast enhancement (ACE)," it consists of limitation in terms of

high processing time (32–38). In MBD classification, local details are more important than global features and reduce processing time; the proposed architecture uses the contrast limited adaptive histogram equalization (CLAHE) algorithm for limited contrast enrichment of fibroglandular tissues (39–41). The first merit of this method helps to minimize the edge shadowing effect and noise produced in homogeneous input digital mammograms. Second, small images known as tiles are used instead of the entire image to perform the CLAHE operation. Hence, contrast enhancement of each tile histogram matches with exponential distribution or Rayleigh distribution. Moreover, to overcome artificial-induced borders, neighboring tiles are connected by bilinear insertion. This advanced CLAHE technique is outlined under:

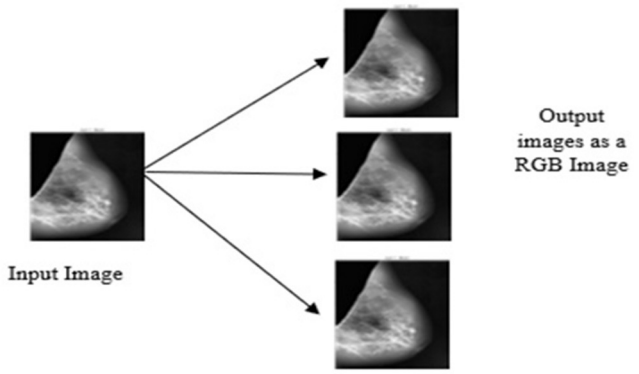


FIGURE 6 | Conversion of grayscale image appears as an RGB.

1. Initially, all the input mammograms are divided into 8 × 8 non-overlapping contextual fields of equal sizes and later a histogram of various contextual regions is calculated.
2. The clip limit (β) is the threshold parameter used to alter the contrast of the image, which is calculated by Equation (1).

$$\beta = \frac{M \times N}{L}\left(1 + \frac{\propto}{100}\left(S_{\max} - 1\right)\right) \quad (1)$$

Where β is the clip termination, calculated as eight by several experiments, M × N is the number of pixels in each field, L is the number of gray scales, and α is a clip factor (0–100). It is the highest allowable slope, which is set to be 4 for this analysis.

3. Each histogram is aligned, so that its maximum does not surpass further than the clip boundary.

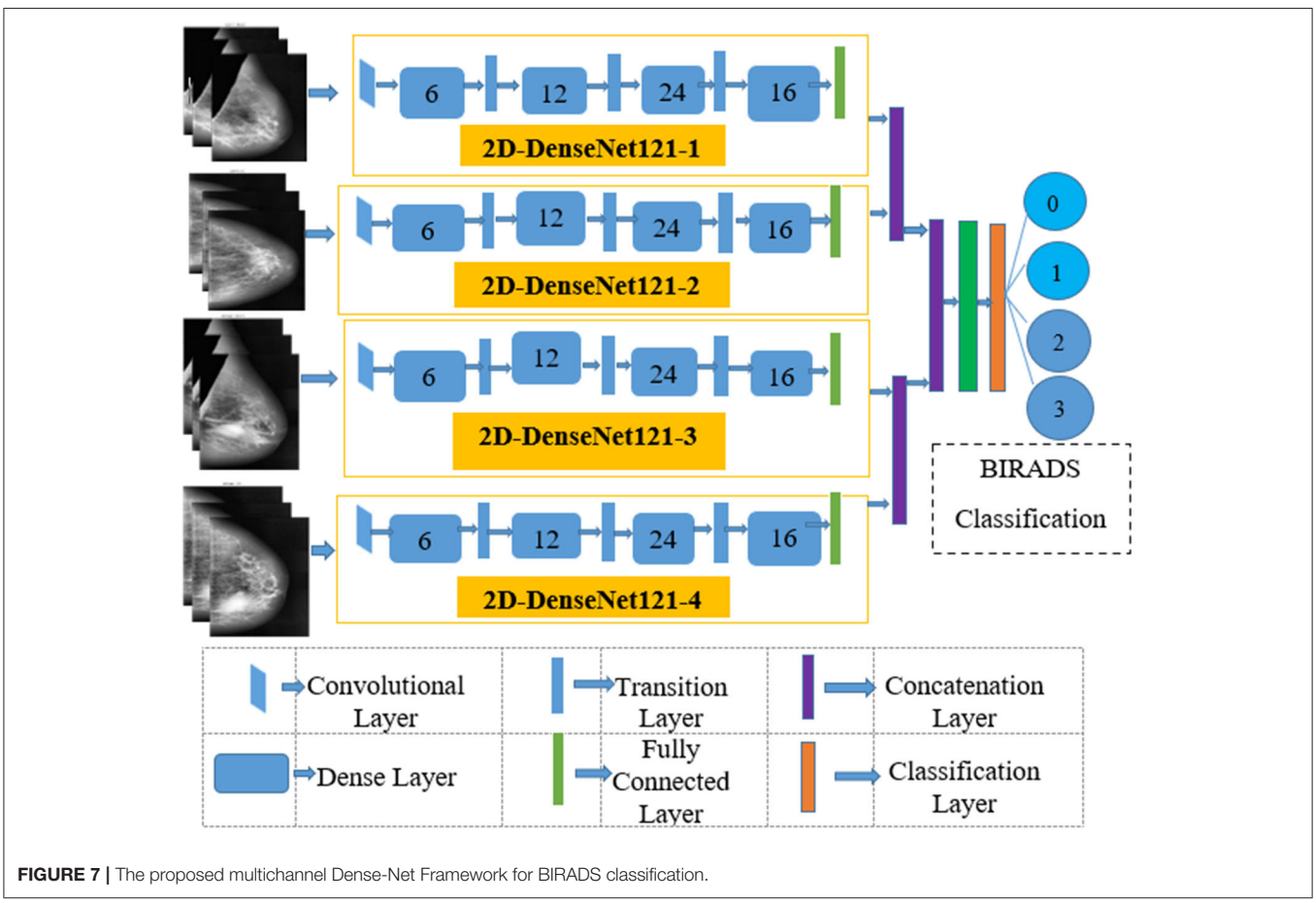


FIGURE 7 | The proposed multichannel Dense-Net Framework for BIRADS classification.

4. The transformation function, which is described below, is used to modify the histogram.

$$t\left(r_k\right) = \sum_{j=0}^{k} p_r\left(r_j\right) \quad (2)$$

$$\text{Where } p_r\left(r_j\right) = \frac{n_j}{n} \quad (3)$$

Equations (2) and (3) describe the probability function of input image gray scale value j, n is a total number of pixels in input mammogram image, and n_j is input pixel number of gray value j.

5. The adjacent tiles were joined by bilinear interpolation and the image gray scale values were altered according to the revised histograms.

In this method, the contrast factor is limited to 0.01 to prevent oversaturation of the image, specifically inhomogeneous areas for optimized output. Furthermore, the number of bins for the histogram structure is restricted to 64 over the uniform distribution for contract enhancing transformations. **Figure 5** depicts the result of contrast enhancement of the CLAHE algorithm on input images.

Multichannel Model Development

This article proposes the feature learning ability of multichannel DenseNet architecture presented by Huang et al. (18) toward MBD classification. The proposed method uses four independent DenseNet architecture as four-channel architecture known as multichannel architecture. This architecture is competent in taking all the four views of an individual patient for the classification of MBD.

Conversion of Gray Scale Image Into RGB

The DenseNet model is pretrained on red, green and blue channel (RGB) images, but the proposed study uses the gray scale image as the input image. To appear gray scale image as an RGB, perform repetition of image array three times due to which the same image appears on the channels. Then, after duplication of the input image, all the input images are resized into 320 × 320 × 3. **Figure 6** depicts the conversion of the gray scale image into a three-channel RGB.

Input Convolutional Layer

The four input channels of the proposed architecture are marked as L_CC, L_MLO, R_CC, and R_MLO. The fundamental merit of this combination is that all the four views of digital mammography are processed concurrently. Each input layer of DenseNet architecture consists of a convolution layer of the kernel of 7 × 7 with a stride of 2. This convolution operation reduces the input size of the images to 112 × 112 × 3. Input image further passes through a pooling layer of 3 × 3 maximum pooling with stride 2 × 2. Thus, the input layer's convolution and pooling operation reduce the input image size to 56 × 56 × 3 and before passing to the dense blocks. **Figure 7** depicts the proposed multichannel architecture.

Design of DenseNet Neural Structure

The DenseNet architecture consists of different design variants such as DenseNet121, DenseNet169, DenseNet 201, and DenseNet 264. The DenseNet architecture's fundamental merit is the structure of dense layers precisely designed to take care of downsampling and feature concatenation. Therefore, out of four variants, the proposed architecture uses the DenseNet121 network structure, consisting of a combination of dense block layers and transition layers. Thus, the proposed model uses 58 convolutional layers and a growth rate ($k = 12$), including four dense and two transition layers. In addition, the proposed model consists of comparatively fewer parameters hence, saving computational memory and reducing the overfitting.

Dense Block Layer

In four dense blocks, the individual layer is responsible for forming a k-characteristic map after convolution, which also maintains feature maps of each layer are in the same size. K convolution kernels extract all the features from the layers. Parameter k is known as a hyperparameter in DenseNet, which is the growth rate of the network. Each dense layer receives the different inputs from previous layers to reduce computation and enhance the efficiency of the dense block. The dense block internally uses the bottleneck layer (1 × 1 convolution layer between batch normalization, ReLU, and 3 × 3 convolution layer).

Transition Layer

This section consists of a batch normalization layer and a one × one convolution layer followed by a two × two average pooling layer. This layer combines two nearby dense block layers to reduce the feature map size. A combination of 4 dense blocks and transition layers converts the image size into 7 × 7 × 3, further provided to the output layer. Each layer connects to the previous stage as an input described by Equation (4):

$$X_l = H_l(\left[x_0, x_1, \ldots\ldots, x_{l-1}\right]) \quad (4)$$

A non-linear transformation function $H_l(.)$ is responsible for combining series output of batch normalization, ReLU, pooling, and convolution operation. **Figure 8** depicts the design architecture of dense layer.

Output Classification Layer

The output layer of the proposed architecture consists of a specific average pooling layer for each channel to extract meaningful features. Extracted features are flattened by the flatten layer and are given to the individual dense layer. MBD features received from four-channel are concatenated together with two concatenation blocks. Subsequently, the third concatenation block joins all the proposed method features together. The three dense layers accept all the features collected together and, finally, the classification layer receives the output of three dense layers for classification. The proposed method uses the SoftMax classifier to classify output into four classes as per the BIRADS Density Classification. **Table 2** presents the specifications of the proposed method.

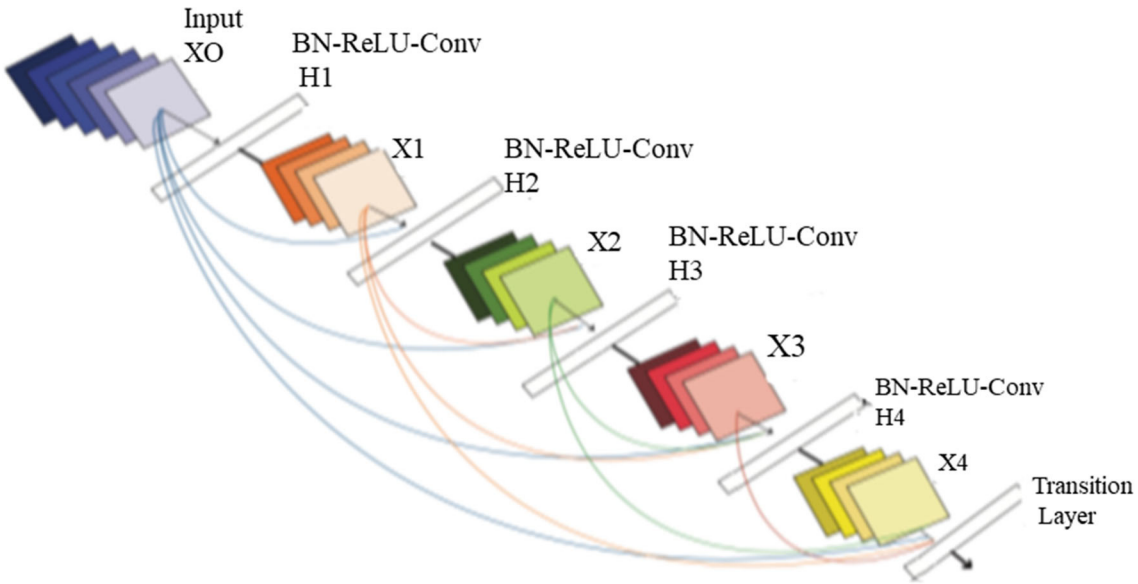


FIGURE 8 | The architecture of dense layer.

ANALYSIS OF EXPERIMENT AND OUTCOMES

The experimentally proposed design is trained and tested upon the PyTorch framework on Google Colaboratory, a free online cloud-based Jupiter notebook environment. The proposed method input dataset is not sufficient to split into the train, validate, and test data sets; hence, training and testing of the model are done in two phases. **Figure 9** depicts the image data distribution during training (phase I) and training and testing (phase II).

Phase I

The entire model is trained with stochastic gradient descent (SGD) algorithm using batch sizes 4 and 30 epoch on the whole dataset. SGD is an optimization algorithm that estimates the error gradient for the model's current state with an example of a training set; after this, it updates the weights of the model using backpropagation (35–37). Equation (5) describes weight updating mechanism in SGD algorithm.

$$w_{new} = w_{old} - n\nabla Qi(w_{new}) \quad (5)$$

Where w_{new} is new weight, is weight at previous iteration is learning rate and weight gradient. The primary merit of this algorithm is that it updates parameters for each training example and performs one update at one time. Thus, SGD is faster and can also learn online. The weight updating step size is the learning rate of the model. The learning rate is a configurable hyperparameter that controls the speed by which the model learns. The initial learning rate for this model is 0.1 (default value) and further divided by ten at 50 and 75% of the total training epochs. The categorical cross-entropy acts as a loss function in this model, quantifying the difference between four probability distributions. This loss function works well with the SoftMax activation function in multiclass classification. Equation (6) describes the categorical cross-entropy mathematically, which is:

$$C.E. = -\sum_{i}^{c} t_i log(s_i) \quad (6)$$

Where $C.E.$ is cross-entropy t_i and s_i ground truth and the convolutional neural network (CNN) score for each class i in c. **Table 3** presents the setting of different hyperparameters used to obtain the optimized results of the proposed architecture.

During the training on the entire dataset, the best classification accuracy score was 96.35% at 18 epochs with a loss factor of 0.1344. **Figure 10** depicts the training phase results on the dataset as a whole.

Phase II

After training the model on the entire dataset, the proposed model performance is validated by spitting the image dataset in a ratio of 80% as training and 20% as testing. The proposed model performed significantly well on all the BIRADS density classes during the testing phase and recorded the best classification accuracy, 90.00%, with a validation loss of 0.3814. **Figure 11** depict the outcomes of validation over the training model.

Results Evaluation

The proposed multichannel DenseNet architecture performance is analyzed from the confusion matrix of the model on the test dataset. **Figure 12A** shows the proposed architecture's heat map (confusion matrix) on the test dataset. The heat map helps to analyze which category is correctly classified by the proposed architecture. The main diagonal darker

TABLE 2 | Technical specification of proposed architecture.

No. of channels	04	
Layers/Channel	**Output size/channel**	**Block description/Channel**
Convolution	112 by 112	Kernel 7 × 7 stride 2
Pooling	56 by 56	3 × 3 Max.Pooling, stride 2 × 2
Dense block 1	56 by 56	[1 × 1 Conv] × 6 [3 ×3 Conv] × 6
Transition 1	56 × 56	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	28 × 28	
Dense block 2	28 × 28	1 × 1 Conv] × 12 [3 × 3 Conv] × 12
Transition 2	28 × 28	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	14 × 14	
Dense block 3	14 × 14	1 × 1 Conv] × 24 [3 × 3 Conv] × 24
Transition layer 3	14 × 14	Batch normalization layer and a 1 × 1 convolution layer followed by 2 × 2 average pooling layer
	7 × 7	
Dense block 4	7 × 7	1 × 1 Conv] × 16 [3 × 3 Conv] × 16
Classification layer	1 ×1	7 × 7 global average pool
	1000D fully connected, SoftMax	

FIGURE 9 | The distribution of image data.

TABLE 3 | Setting of hyperparameters used during experiment.

Hyper parameters	Value
Model	Multichannel-Dense Net
No. of channel	04
Model initial learning rate	0.1
Image size	320 × 320 × 3
Batch size	04
Target labels	Ground Truth
Data augmentation	Not used
Loss function	Categorical cross-entropy
Optimization algorithm	Stochastic gradient decent
Validation parameter	Classification accuracy

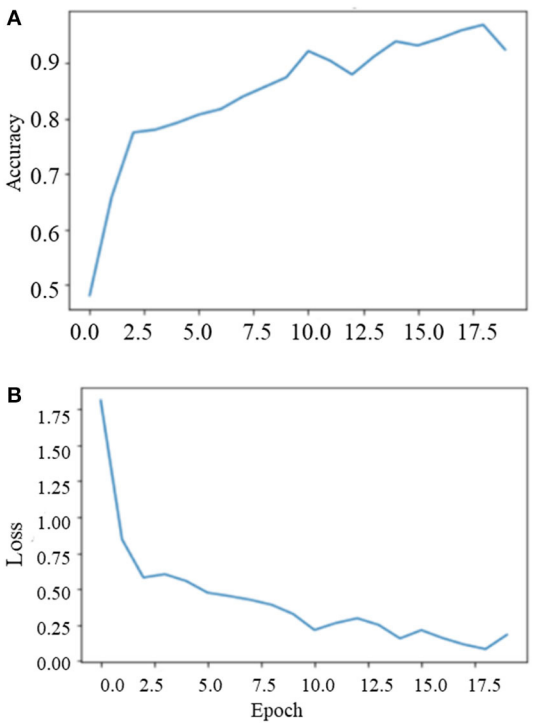


FIGURE 10 | Training phase performance of the model **(A)** model accuracy and **(B)** model loss.

version indicates a better classification rate. Although it is clear that the model is working well in classes A, C, and D, there is some confusion in classifying category class B. Still, this model correctly classifies heterogeneous dense (C) and extremely dense (D), which is the essential bottleneck behind MBD classification. **Figure 12** depict the heat map and the AUC curve, respectively, of the proposed model.

Evaluation of the classification results of the intended architecture is performed in terms of precision, recall, the F1-score, and classification accuracy. Among those parameters, precision is the proportion of samples with optimistic predictions concerning the total number of correct positive samples. The recall ratio of correctly predicted samples to the whole samples and the F1-score are the precision and recall weight. Finally, classification accuracy is the total correct predictions to the total number of samples. Equations (7) to (10) define precision, recall, the F1-score, and classification accuracy, respectively:

$$Acc = \frac{N_A + N_B + N_C + N_C}{N_{All}} \tag{7}$$

$$F1 - score = \frac{2PR}{P + R} \tag{8}$$

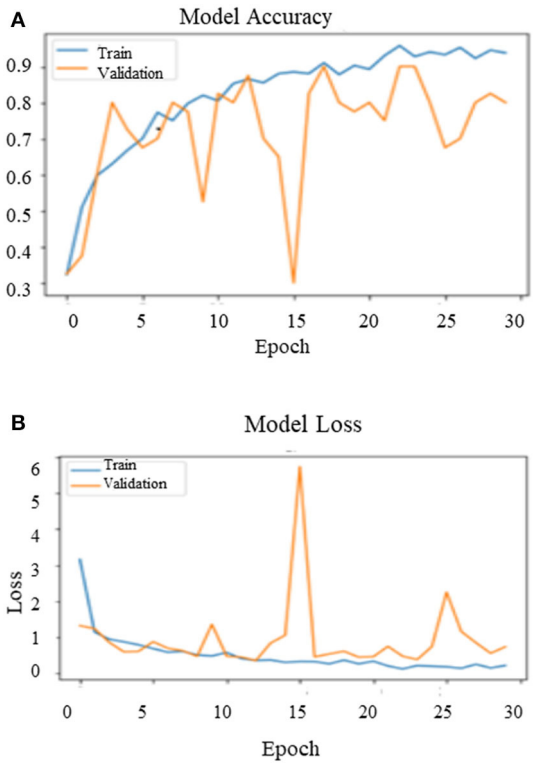


FIGURE 11 | Validation results of the proposed model in phase-II **(A)** model accuracy **(B)** model loss.

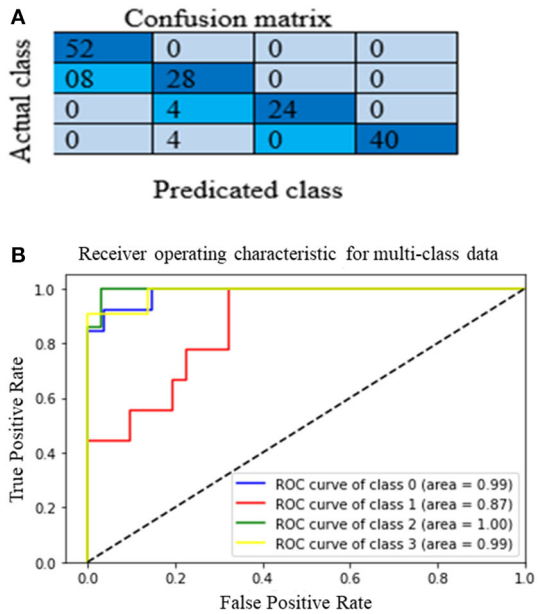


FIGURE 12 | (A) The Heat map **(B)** and the ROC curve of the proposed model.

Where N_{All} *is* the total number of images and *Na*, *Nb*, *Nc, and Nd* signify the number of images in MBD classes A, B, C, and D.

$$P = \frac{N_t}{N_i} \times 100 \tag{9}$$

$$R = \frac{Nt}{Nf} \times 100 \tag{10}$$

Where, Nt is the correct number of predictions of a specific category and Ni is all the number of forecasts of a class and indicates the actual number of the category, among them are precision and recall, respectively. The model accuracy is the ratio of the sum of the diagonal elements to all the elements. Thus, it acts as an indicator of the overall prediction of the model. **Table 4** presents the model's overall performance in detail and variation in precision rate, the recall rate, and the F1-score rate under different categories. The number (or percentage) of accurate positive samples for all the four BIRADS density classes are 92, 75.5, 92.2, and 94.7% of their respective totals. From the results shown in **Table 4**, there is no confusion between classes A and C, B and C, and C and D. The proposed algorithm results are consistent with the results evaluated by the radiologists, which are a positive sign that indicates that deep learning models are helpful for the classification of MBD. Another graphical technique utilized to investigate the realization of computer-aided diagnostic methods is the receiver operating characteristic (ROC), as shown in **Figure 12B**. This curve analysis performance of the prototype should be in terms of true positive rate (TP) and false positive rate (FP). The AUC value refers to the area enclosed by the ROC curve in the [0, 1] period and the X-axis. The higher the AUC content, the more reliable the realization of the model. It additionally highlights the ability of the model to differentiate among the classes.

DISCUSSION

Inconsistency in MBD assessment is the fundamental reason behind unnecessary extra screening procedures and cause for patient anxiety (41–43). Due to improved system architecture and hardware capability, the deep learning model can be an alternative for medical image classification. Still, the need for a larger dataset and vanishing gradient are the primary bottleneck issues to obtain state-of-the art results from deep learning models (44).

Advantages of Proposed Method

This study article proposes multichannel DenseNet architecture for MBD classification to investigate the performance of DenseNet architecture on a smaller dataset. The proposed architecture has recorded good classification performance in four-class classification. Furthermore, the visualization results show that the model can distinguish between all the BIRADS density classes, especially in "scattered density" and "heterogeneously dense category of the BIRADS." Thus, this model can help the radiologists to classify the BIRADS density classes quickly. The main reasons for the excellent performance of this model are as follows:

TABLE 4 | Performance parameter of the proposed method.

BIRADS density classes	Precision	Recall	F1-score	Overall classification accuracy	Overall AUC
Predominantly fatty-class A	1	0.866	0.92	0.9006	0.9625
Fat with some fibro glandular tissue class B	0.77	0.77	0.755		
Heterogeneous dense-class C	0.857	1	0.922		
Extremely dense-class D	0.90	1	0.947		

TABLE 5 | Comparative status of the proposed method with current state-of-the-art methods.

References	Dataset	Proposed method	Classification accuracy
Wu et al. (20)	2,00,000	A deep convolutional network with 100 layers.	0.825 on Four views
Ciritsis et al. (19)	20,578	A deep convolutional network with 11 layers and performed analysis separately on CC and MLO views.	0.897 On CC views and 0.866 on MLO views.
Kaiser et al. (24)	8,150	A multichannel architecture with transfer learning by VGG-Net.	0.88 on all four views
Shi et al. (22)	322	A light-weight deep learning architecture with 3 convolutional layers.	0.836 On MLO views.
Deng et al. (36)	18,157	A single channel architecture with transfer learning by Dense Net 121 combined with SE-Attention network.	0.9179 on all Four views
Proposed method	800	A multichannel architecture with transfer learning with Dense Net 121	0.90 on Four views

1. Instead of using raw images, the proposed architecture uses preprocessed images. Hence, there are no high-density areas such as pectoral muscle and tags on mammograms, which are helpful to increase the classification accuracy of the proposed architecture.
2. The proposed method uses a contrast enhancement technique to improve the quality of training data.
3. The model contains four DenseNet branches, which extract the features of mammograms from four views of the single patient, so that the network can focus on a broader range of spatial information.
4. Due to multichannel architecture, it is possible to process all the views of a single patient simultaneously.
5. Subsequently, with multichannel and multiview architecture, it is probable to consolidate all the features collectively; consequently, the performance of the intended model is found more reliable than single-view classification.

Comparison With Existing Methods

While there are variations in related datasets and evaluation methods, the straightforward comparison is challenging to researchers. This section compiles the proposed algorithm's comparison state with existing classifications. To study the interobserver variation in MBD assessment, N Kaiser et al. (24) proposed the novel multichannel VGG architecture. This approach uses a total of 8,150 digital mammograms, divided into 600 cases. This method recorded 88% two-class classification accuracy (dense and nondense) with the AUC of 0.954. Besides, these results are also compared with the 32 individual radiologist's panel's density ground truth. This study reveals that the deep learning approach performs better than average radiologists. Thus, we only need to refine the deep learning model for MBD classification. However, the fundamental limitation of this method is that the gradient flows from the final layer to the initial layer; hence, vanishing gradient problem takes place, which increases training time and reduces the classification accuracy. In the proposed method, due to DenseNet architecture, all the layers are directly connected in feed-forward nature, acting as an effective solution for vanishing gradient and reducing training time. Thus, the results of the proposed algorithm outperform this method on a smaller dataset.

Another method directly comparable to the proposed method is the optimized lightweight deep learning architecture proposed by Shi et al. (22). The elemental focus of this method is to overcome the requirement of the larger dataset and vanishing gradient problem of the deep learning algorithm. This method combines three CNN layers, one dense layer, and an output layer to classify MBD. This architecture is tested on the 322 mini-MIAS dataset with different data augmentation techniques and recorded 83.6% classification accuracy. However, due to the smaller dataset, this architecture has limitations regarding moderate classification accuracy and low stability of the network. Therefore, the proposed method used the concept of transfer learning and multichannel architecture to overcome these limitations. As a result, the proposed model outperforms this method in classification accuracy on a smaller dataset. **Table 5** provides the comparative state of the proposed method with other different existing methods.

Limitations and Future Study

Even though the proposed method has improved the classification performance of the BIRADS density classes, some issues still need to be addressed. First, this study uses a smaller amount of image data and no image enhancement strategies are used to expand the dataset. Hence, model performance, especially stability during validation, is affected

due to limited image data and model found it a little confusing to classify classes A and B. Therefore, in future study, data enhancement techniques will improve the model's performance. Second, the proposed study addresses only one type of dataset; hence, this approach does not address the robustness of the model. Future study will address the robustness of the model by training the model with different vendor-specific image datasets and testing results of all the mentioned state-of-the-art methods with the proposed method with the same work environment. The proposed study will be undoubtedly helpful in addressing the issues mentioned above.

CONCLUSION

In summary, the primary objective behind this study is to classify MBD as per the BIRADS classification. This study proposes the novel approach of multichannel architecture with DenseNet121 for the objective assessment of MBD. The proposed framework uses the four views of a single patient to enhance feature learning ability through a multiview approach. In the method, image contrast enhancement and preprocessing of the input image are implemented to enhance the condition of the training image data. The input images are processed through multichannel architecture to extract and fuse all the features. Analysis of the results suggests that the proposed model successfully distinguishes between all the BIRADS density classes, but is predominantly found superior in the two most distinctive and challenging BIRADS categories: "BIRADS_C" and "BIRADS_D." Classification accuracy of the proposed model is recorded at 96.67% during training and 90.06% during testing and the average AUC of 0.9625. The introduced design consists of some weaknesses discussed and will be addressed in future study; with certain modifications, the proposed method is suitable for application in clinical workflow in breast cancer screening to avoid false recalls.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

The authors are grateful to the Taif University Researchers Supporting Project number (TURSP-2020/36), Taif University, Taif, Saudi Arabia and gratefully acknowledge Dr. Yogesh G. Yadav and Dr. Atul E. Sawant, Specialist Radiologist, Lifecare Hospital, Musaffah, Abu Dhabi for their valuable and timely help in marking ground truth on some difficult cases of mammogram from DDSM dataset and providing subjective evaluation of same.

REFERENCES

1. Advani P, Moreno-Aspitia A. Current strategies for the prevention of breast cancer. *Breast Cancer Targets Therapy*. (2014) 6:59–71. doi: 10.2147/BCTT.S39114
2. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol.* (2006) 24:2137–50. doi: 10.1200/JCO.2005.05.2308
3. Sapate S, Talbar S, Mahajan A, Sable N, Desai S, Thakur M. Breast cancer diagnosis using abnormalities on ipsilateral views of digital mammograms. *Biocybern Biomed Eng.* (2020) 40:290–305. doi: 10.1016/j.bbe.2019.04.008
4. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee J, Pasupulet S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020). 8:218–25. doi: 10.3389/fpubh.2020.00357
5. Dhanamjayulu C, Nizhal U, Maddikunta P, Gadekallu T, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process.* (2021) 16:647–58. doi: 10.1049/ipr2.12222
6. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia System*. (2021) 1:1–15. doi: 10.1007/s00530-021-00774-w
7. Ngabo D, Dong W, Ibeke E, Iwendi C, Masabo E. Tackling pandemics in smart cities using machine learning architecture. *Math Biosci Eng.* (2021) 18:8444–61. doi: 10.3934/mbe.2021418
8. Dungan JS. Mammographic density and the risk and detection of breast cancer. *Obstetr Gynecol Women Health*. (2008) 1:214–5. doi: 10.1016/S1090-798X(08)79014-3
9. Lam PB, Vacek PM, Geller BM, Muss HB. The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer*. (2000) 89:369–75. doi: 10.1002/1097-0142(20000715)89:2<369::aid-cncr23>3.0.co;2-j
10. Sapate SG, Mahajan A, Talbar SN, Sable N, Desai S, Thakur M. Radiomics based detection and characterization of suspicious lesions on full field digital mammograms. *Comput Methods Programs Biomed.* (2018) 163:1–20. doi: 10.1016/j.cmpb.2018.05.017
11. Rice MS, Bertrand KA, Lajous M, Tamimi RM, Torres G, López-Ridaura R, et al. Reproductive and lifestyle risk factors and mammographic density in Mexican women. *Ann Epidemiol*. (2015) 25:868–73. doi: 10.1016/j.annepidem.2015.08.006
12. Pawar S, Sapate S, Sharma K. Machine learning approach towards mammographic breast density measurement for breast cancer risk prediction: an overview. *Proc ICAST*. (2020) 2020:1–14. doi: 10.2139/ssrn.3599187
13. Alyami H, Alosaimi W, Krichen M, Alroobaea R. Monitoring social distancing using artificial intelligence for fighting COVID-19 virus spread. *Int J Open Source Softw Proc.* (2021) 12:48–63. doi: 10.4018/IJOSSP.2021070104
14. Arul R, Alroobaea R, Tariq U, Almulihi AH, Alharithi FS, Shoaib U. IoT-enabled healthcare systems using block chain-dependent adaptable services. *Pers Ubiquitous Comput.* (2021) 1:1–15. doi: 10.1007/s00779-021-01584-7
15. Devi SS, Vidivelli S. Classification of breast tissue density in digital mammograms. In: *Proceedings of ICIIECS*. Coimbatore (2018). p. 1–7. doi: 10.1109/ICIIECS.2017.8276139
16. Srivastava RK, Greff K, Schmidhuber J. *Training Very Deep Networks. Advances in Neural Information Processing Systems.* (2015). p. 2377–85.
17. Lee J, Nishikawa RM. Automated mammographic breast density estimation using a fully convolutional network. *Med Physics*. (2018) 45:1178–90. doi: 10.1002/mp.12763

18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *Proceedings of CVPR*. Los Alamitos, CA (2017). p. 2261–9. doi: 10.1109/CVPR.2017.243
19. Ciritsis A, Rossi C, De Martini I, Eberhard M, Marcon M, Becker AS, et al. Determination of mammographic breast density using a deep convolutional neural network, *Br J Radiol*. (2019) 92:20180691. doi: 10.1259/bjr.20180691
20. Wu N, Geras K, Shen Y, Su J, Kim SG, Kim E, et al. Breast density classification with deep convolutional neural networks. In *Proceedings of ICASSP*. Calgary, AB (2018). p. 6682–6. doi: 10.1109/ICASSP.2018.8462671
21. Lizzi F, Laruina F, Oliva P, Retico A, Fantacci ME. Residual convolutional neural networks to automatically extract significant breast density features. *Commun Comput Inform Sci*. (2019) 1089:28–35. doi: 10.1007/978-3-030-29930-9_3
22. Shi P, Wu C, Zhong J, Wang H. Deep learning from small dataset for bi-rads density classification of mammography images. In: *Proceedings of ITME*. Qingdao (2019). p. 102–9. doi: 10.1109/ITME.2019.00034
23. Bengio Y. Learning deep architectures for AI. *Found Trends Mach Learn*. (2009) 2:1–27. doi: 10.1561/2200000006
24. Kaiser N, Fieselmann A, Ritschl L, Kappler S, Ravikumar N, Ritschl L, et al. Mammographic breast density classification using a deep neural network: assessment on the basis of inter-observer variability. *SPIE-Intl Soc Opt Eng*. (2019) 1:23. doi: 10.1117/12.2513420
25. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer WP. *International workshop on digital mammography*. Toronto, ON: *Med Physics* in *Proceedings of IWDM* (2001).
26. Sapate S, Talbar S. An overview of pectoral muscle extraction algorithms applied to digital mammograms. *Stud Comput Intell*. (2016) 651:19–54. doi: 10.1007/978-3-319-33793-7_2
27. Pawar SD, Sharma KK, Sapate SG. *Design of Intelligent Applications Using Machine Learning and Deep Learning Techniques* 1st ed. London: Francis and Taylor; CRC Press. (2021). p. 125–43. doi: 10.1201/9781003133681-8
28. Pawar SD, Sharma KK, Sapate SG, Yadav GY. Segmentation of pectoral muscle from digital mammograms with depth-first search algorithm towards breast density classification. *Biocy Biomed Eng*. (2021) 41:1224–41. doi: 10.1016/j.bbe.2021.08.005
29. Zeng M, Li Y, Meng Q, Yang T, Liu J. Improving histogram-based image contrast enhancement using gray-level information histogram with application to X-ray images. *Optik*. (2012) 123:511–20. doi: 10.1016/j.ijleo.2011.05.017
30. Öktem H, Egiazarian K, Niittylahti J, Lemmetti J. An approach to adaptive enhancement of diagnostic X-Ray images. *Eur J Appl Sig Process*. (2003) 5:430–6. doi: 10.1155/S1110865703211069
31. Deng G. A generalized unsharp masking algorithm. *IEEE Transac Image Process*. (2011) 20:1249–61. doi: 10.1109/TIP.2010.2092441
32. Chang DC, Wu WR. Image contrast enhancement based on a histogram transformation of local standard deviation. *IEEE Transac Med Imaging*. (1998) 17:518–31. doi: 10.1109/42.730397
33. Bourouis S, Channoufi I, Alroobaea R, Rubaiee S, Andejany M, Bouguila N. Color object segmentation and tracking using flexible statistical model and level-set. *Multimedia Tool Appl*. (2021) 80:5809–31. doi: 10.1007/s11042-020-09809-2
34. Hedabou M. *Computer and Cyber Security: Principles, Algorithm, Applications and Perspective* 1st ed. London: Francis and Taylor Publisher; CRC Press (2018).
35. Toulis P, Airoldi EM. Asymptotic and finite-sample properties of estimators based on stochastic gradients. *Ann Stat*. (2017) 45:1694–727. doi: 10.1214/16-AOS1506
36. Deng J, Ma Y, Li D-A, Zhao J, Liu Y, Zhang H. Classification of breast density categories based on SE-Attention neural networks. *Comput Method Program Biomed*. (2020) 193:105489. doi: 10.1016/j.cmpb.2020.105489
37. Iggaramen Z, Hedabou M. FADETPM: Novel approach of file assured deletion based on trusted platform module in lecture notes in networks and systems. *Springer Verlag*. (2017) 49:49–59. doi: 10.1007/978-3-319-97719-5_4
38. Azougaghe A, Hedabou M, Belkasmi M. An electronic voting system based on homomorphic encryption and prime numbers. In: *Proceedings of ICIAS*. Kuala Lumpur (2016). p. 140–5. doi: 10.1109/ISIAS.2015.7492759
39. Bentajer A, Hedabou M, Abouelmehdi K, Igarramen Z, El Fezazi S. An IBE-based design for assured deletion in cloud storage. *Cryptologia*. (2019) 43:254–65. doi: 10.1080/01611194.2018.1549123
40. Mukhtar H, Rubaiee S, Krichen M, Alroobaea R. An Iot framework for screening of covid-19 using real-time data from wearable sensors. *Int J Environ Res Public Health*. (2021) 18:56–62. doi: 10.3390/ijerph18084022
41. Mubashar A, Asghar K, Javed AR, Rizwan M, Srivastava G, Gadekallu TR, et al. Storage and proximity management for centralized personal health records using an ipfs-based optimization algorithm. *J Circuits Syst Comput*. (2022) 31:2250010. doi: 10.1142/S0218126622500104
42. Arikumar KS, Prathiba SB, Alazab M, Gadekallu TR, Pandya S, Khan JM, et al. FL-PMI: federated learning-based person movement identification through wearable devices in smart healthcare systems. *Sensors*. (2022) 22:1377. doi: 10.3390/s22041377
43. Rajput DS, Basha SM, Xin Q, Gadekallu TR, Kaluri R, Lakshmanna K, et al. Providing diagnosis on diabetes using cloud computing environment to the people living in rural areas of India. *J Ambient Intell Humaniz Comput*. (2021) 10:1–12. doi: 10.1007/s12652-021-03154-4
44. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *PeerJ Comput Sci*. (2021) 7:e390. doi: 10.7717/peerj-cs.390

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Pawar, Sharma, Sapate, Yadav, Alroobaea, Alzahrani and Hedabou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fpubh.2022.892371

Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease

Michael Onyema Edeh[1], Surjeet Dalal[2], Imed Ben Dhaou[3]*, Charles Chuka Agubosim[4], Chukwudum Collins Umoke[5], Nneka Ernestina Richard-Nnabu[6] and Neeraj Dahiya[7]*

[1] Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria, [2] College of Computing Science & Information Technology, Teerthanker Mahaveer University, Moradabad, India, [3] Department of Computer Science, Hekma School of Engineering, Computing and Informatics, Dar Al-Hekma University, Jeddah, Saudi Arabia, [4] Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University, Uli, Nigeria, [5] Department of Vocational and Technical Education, Alex Ekwueme Federal University Ndufu Alike Ikwo (AE- FUNAI), Abakaliki, Nigeria, [6] Department of Computer Science/Informatics, Alex Ekwueme Federal University Ndufu Alike Ikwo (AE-FUNAI), Abakaliki, Nigeria, [7] Department of Computer Science and Engineering, SRM University Delhi-NCR, Sonipat, India

Machine learning algorithms are excellent techniques to develop prediction models to enhance response and efficiency in the health sector. It is the greatest approach to avoid the spread of hepatitis C, especially injecting drugs, is to avoid these behaviors. Treatments for hepatitis C can cure most patients within 8 to 12 weeks, so being tested is critical. After examining multiple types of machine learning approaches to construct the classification models, we built an AI-based ensemble model for predicting Hepatitis C disease in patients with the capacity to predict advanced fibrosis by integrating clinical data and blood biomarkers. The dataset included a variety of factors related to Hepatitis C disease. The training data set was subjected to three machine-learning approaches and the validated data was then used to evaluate the ensemble learning-based prediction model. The results demonstrated that the proposed ensemble learning model has been observed ad more accurate compared to the existing Machine learning algorithms. The Multi-layer perceptron (MLP) technique was the most precise learning approach (94.1% accuracy). The Bayesian network was the second-most accurate learning algorithm (94.47% accuracy). The accuracy improved to the level of 95.59%. Hepatitis C has a significant frequency globally, and the disease's development can result in irreparable damage to the liver, as well as death. As a result, utilizing AI-based ensemble learning model for its prediction is advantageous in curbing the risks and improving treatment outcome. The study demonstrated that the use of ensemble model presents more precision or accuracy in predicting Hepatitis C disease instead of using individual algorithms. It also shows how an AI-based ensemble model could be used to diagnose Hepatitis C disease with greater accuracy.

Keywords: artificial intelligence, machine learning, hepatitis C, Quest, ensemble learning

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Varun Malik, Chitkara University, India
Amena Mahmoud, Kafrelsheikh University, Egypt

****Correspondence:***
Michael Onyema Edeh
michael.edeh@ccu.edu.ng
Imed Ben Dhaou
imed.bendhaou@utu.fi

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 March 2022*
Accepted: *17 March 2022*
Published: *27 April 2022*

Citation:
Edeh MO, Dalal S, Dhaou IB, Agubosim CC, Umoke CC, Richard-Nnabu NE and Dahiya N (2022) Artificial Intelligence-Based Ensemble Learning Model for Prediction of Hepatitis C Disease. Front. Public Health 10:892371. doi: 10.3389/fpubh.2022.892371

INTRODUCTION

Hepatitis C is a liver condition caused by the hepatitis C virus (HCV), which affects the immune system. The main reason behind the spreading of Hepatitis C is through the blood contact of an infected person. Life-threatening and serious health issues such as Cirrhosis and liver cancer are the result of chronic hepatitis C. Often, the early-stage symptoms or the sickness don't appear in

people suffering from chronic hepatitis C. When these symptoms appear in patients, they might often show signs of advanced liver disease (1). Unfortunately, there is still no vaccine available in medical facilities to treat Hepatitis C. By not sharing needles and refraining from behaviors that could lead to the spread of this disease, such as injecting drugs, the most effective strategy to stop HIV from spreading is to avoid it. Treatments for Hepatitis C that have been evaluated and found effective can cure a patient infected with the virus within 8 to 12 weeks if administered on time.

Hepatitis C infections may be considered as acute (short-term) or chronic (long-lasting). If the symptoms last for 6 months, then that patient is considered to be suffering from acute hepatitis. Unfortunately, in more than 50% of the cases, the patient's body becomes unable to clear the virus and hence this acute infection transforms into a chronic disease. According to some researchers' projections, the United States will be free of this chronic Hepatitis C infection by the year 2036. When it comes to better understanding a condition and its prognosis, medical practitioners can use their patient's electronic health information to uncover new findings and trends that might otherwise go unnoticed by physicians. Machine learning classifiers are used in this work to predict the diagnoses of healthy controls and patients who have been diagnosed with hepatitis C. Other viral infections and cancers may benefit from the use of genetic analysis and machine-learning techniques to choose the best course of treatment. The hepatitis C virus can cause substantial liver damage when it infects the liver and produces inflammation. Blood transfusions, injectable drug usage, and sexual behaviors that expose people to blood are the most common ways in which this virus is spread. It is a bloodborne virus.

The acute phase of hepatitis C infection is the first stage of a long-term infection. Hepatitis C is often misdiagnosed due to the lack of symptoms associated with acute hepatitis C infection. Within 1 to 3 months of infection, acute symptoms occur and continue anywhere from 2 to 3 months. It's not usually the case that an acute hepatitis C infection becomes chronic.

Hepatitis C spreads through contact with the blood of someone who has HCV. This contact may be through (2):

- Hepatitis C spreads through sexual transmission.
- Assisting someone who has HIV by sharing drug needles or other drug paraphernalia Hepatitis C is most commonly acquired this way in the United States.
- Getting poked by a needle that was previously used on someone with HCV. The same thing can happen in a hospital or medical facility.
- Using instruments or inks that have not been disinfected after being used on someone with HCV to get tattooed or pierced
- HCV-infected blood or open sores can spread the disease.
- Razor blades and toothbrushes that may have come into contact with someone else's blood should not be shared.
- Being born to a mother infected with the human papillomavirus (HPV).
- Sexual contact with an HCV-infected person that is not protected

In the period before 1992, the common causes of the spread of Hepatitis C were through blood transfusions and transplants of organs. Infections spread by these ways are now extremely rare in the United States, thanks to the widespread use of routine blood supply testing (3). The medical experts have highly recommended the screening of hepatitis C even for people not showing any symptoms of this disease as after analyzing the population the researchers have concluded that people suffering from chronic hepatitis C may doesn't show any symptoms until it causes complications (4).

Artificial intelligence (AI) advancements assist clinicians in providing patients with faster diagnoses and more effective therapy. Artificial intelligence (AI) and human diagnostic efficiency have been compared in studies. AI was found to be just as good at diagnosing as humans, and in some cases better, when compared to less-experienced doctors. In this article, we introduce an artificial intelligence technique based on previously acquired data from a large number of patients (5). The main objectives of this study are given as below:

- O1: Develop a novel Ensemble model of Machine learning algorithms for Hepatitis C prediction.
- O2: Predicate the level of Hepatitis C using machine learning where it plays a significant role in accurately predicting the disease in different age groups.
- O3: Implement Ensemble model to Hepatitis C prediction and compare it to existing model.

The objective O1 is being designed for predicting Hepatitis C. Moreover, for finding the accurate prediction of Hepatitis C, the objective O2 is planned. Objective O3 will be fulfilled by applying the Ensemble model to predict Hepatitis C. The results are then compared with existing Machine Learning algorithms. The main output of these objectives is predicting Hepatitis C disease more accurately.

The remainder of this paper is organized as follows: Section Related Work reviews the related work in the body of literature. Section Materials and Methods discusses the data and materials used followed by section Methods which explains the theory behind the algorithms. Section Results and Discussion presents and analyzes the experimental results. Finally, the paper concludes and proposes future work in section Conclusion.

RELATED WORK

To capitalize on this hepatitis information base, scientists utilized a fleeting reflection technique. We offer novel ideas and strategies for abstracting momentary changed and long haul changed tests in light of hepatitis hidden data and information investigation. Information deliberation empowers us to utilize a wide scope of AI methods to distinguish new and intriguing data for clinical doctors (6–9).

Utilizing AI draws near, Mahmoud ElHefnawi et al. (10) fostered an impressively precise prescient model for Egyptian patients' reactions in view of their clinical and biochemical information. The CART characterization calculation was used for choice trees (DTs). Pruning levels of 9, 11, 13, and 17 were

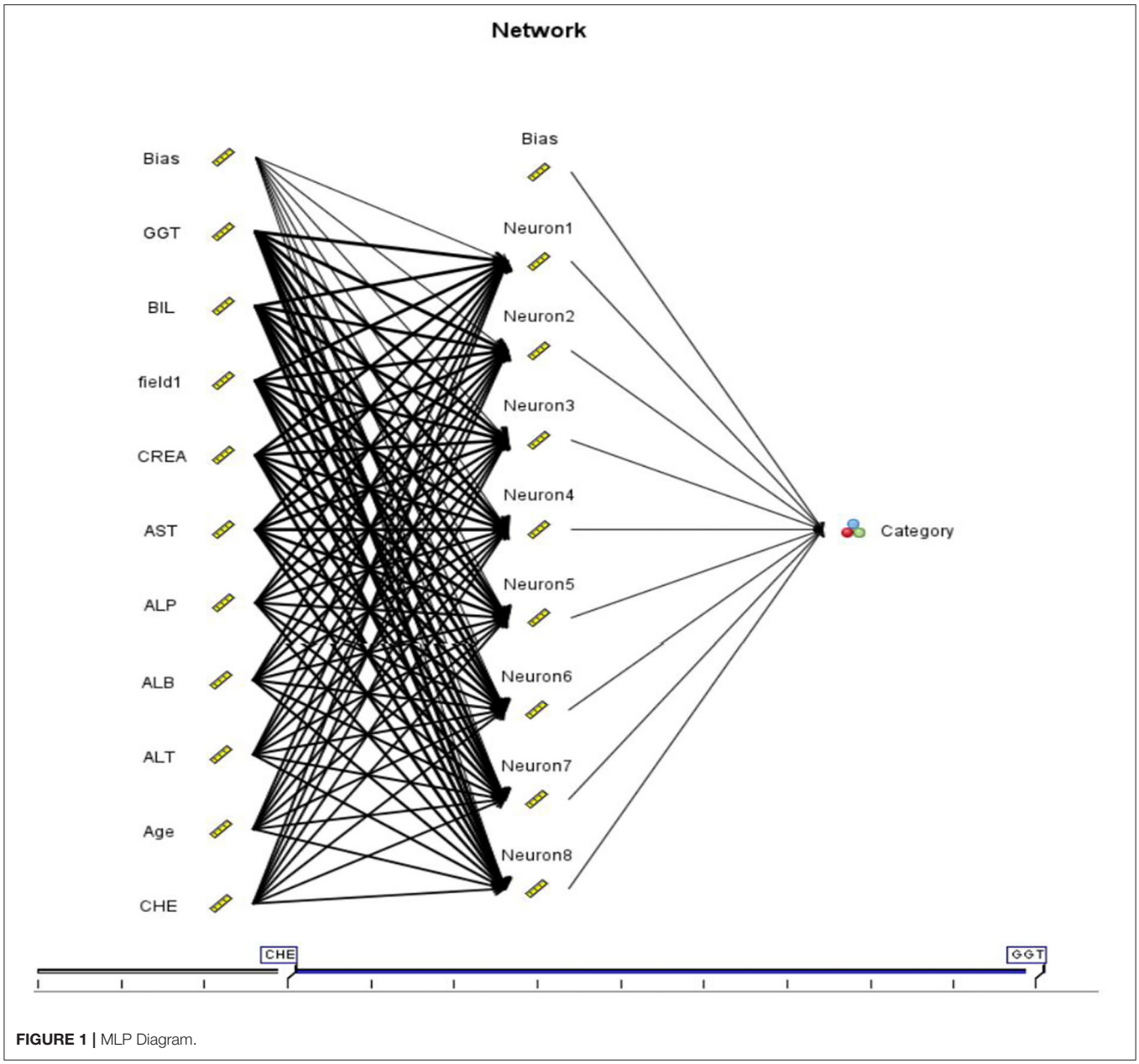


FIGURE 1 | MLP Diagram.

examined, as were hubs going in number from 45 to 61 in every one of the six DTs. Hematology action record and fibrosis, viral burden and Alfa-feta protein and egg whites were the most measurably huge instructive parts of the 12 elements in this examination. At last, a 20% test set was utilized to confirm the models. The ANN and DT models have the most noteworthy and middle exactness of 0.76 and 0.69, separately, and 0.80 and 0.72. This test had an exactness of 0.95 percent, and an explicitness of 0.89 and 0.89%. We've reached the resolution that choice trees are more exact at anticipating reaction and can be utilized to direct patients toward the most ideal course of treatment choices.

The essential justification behind liver transplantation in Latin America is hepatitis C disease. Long haul results, for example, demise are further developed when hepatitis C is treated in contaminated patients. Almost 50% of patients treated with pegylated interferon and ribavirin had a drawn out viral reaction. Over the most recent couple of years, new meds have been fostered that have drastically worked on long haul viral reaction, making triple treatment the norm of care. People with chronic hepatitis C now have a place to go in Latin America thanks to these guidelines.

Using a virtual screening (VS) strategy, Wei et al. (11) discovered new HCV NS5B polymerase inhibitors using a combination of irregular backwoods (RB-VS), e-pharmacophore (PB-VS), and docking techniques. A sum of five mixtures were picked for extra enemy of HCV movement testing and cell cytotoxicity examines from the last hit list. Hindrance of NS5B polymerase by each of the five mixtures was viewed as at IC50

Target	Category
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	8

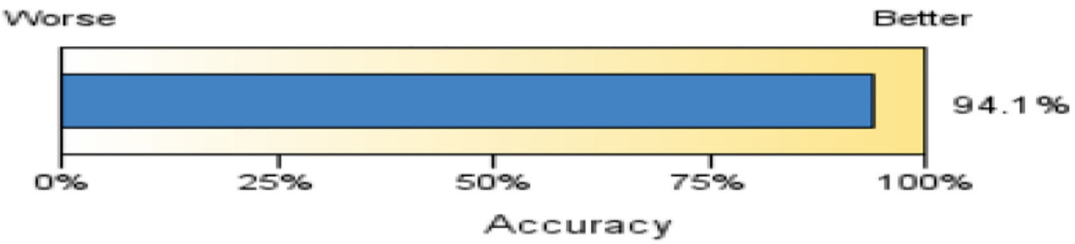


FIGURE 2 | MLP Accuracy.

Classification for Category

Overall Percent Correct = 98.3%

Observed	Predicted				
	0=Blood Donor	0s=suspect Blood Donor	1=Hepatitis	2=Fibrosis	3=Cirrhosis
0=Blood Donor	99.8%	0.0%	0.0%	0.0%	0.2%
0s=suspect Blood Donor	0.0%	100.0%	0.0%	0.0%	0.0%
1=Hepatitis	0.0%	0.0%	90.0%	5.0%	5.0%
2=Fibrosis	0.0%	0.0%	50.0%	41.7%	8.3%
3=Cirrhosis	0.0%	0.0%	0.0%	0.0%	100.0%

Row Percent
100.00
80.00
60.00
40.00
20.00
0.00

FIGURE 3 | MLP Classification for Category.

values going from 2.01 to 23.84 M, and hostile to HCV action at EC50 values going from 1.61 to 21.88 M. All compounds, except for compound N2, showed no cell cytotoxicity (CC50 > 100 M), albeit compound N2 showed feeble cytotoxicity at a CC50 worth of 51.3 M. There was a particular file of 32.1 for N2's HCV antiviral action, making it the most intense. NS5B polymerase inhibitors could be created from the five hit compounds with novel platforms that have been distinguished.

Hepatitis C infection (HCV) habitually shapes a constant contamination, which is regularly asymptomatic in the beginning phases of the illness. There are as of now no symptomatic models for deciding if a HCV contamination is ongoing

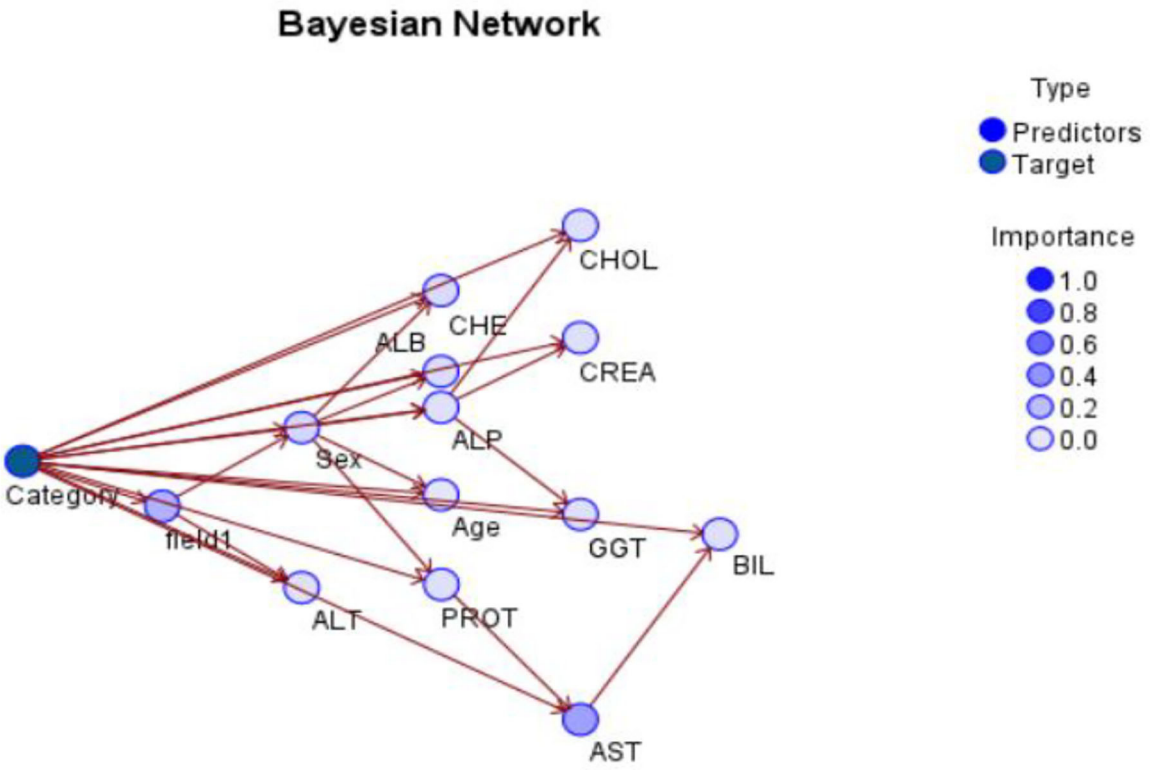


FIGURE 4 | Bayesian network for current problem.

vs. persistent. In light of HCV's inclination to mistake in replication, every tolerant has an assorted assortment of hereditarily unmistakable HCV strains. Therefore, it is for the most part accepted that the degree of intra-have HCV heterogeneity develops over the long haul. Because of elements, for ex-ample, particular breadths and negative choice that happen during persistent disease (2, 12), fundamental measurements for evaluating hereditary heterogeneity are not ex-act enough for HCV contamination organizing in light of the intricacy of the primary improvement of HCV populaces inside has.

In a revelation dataset ($n = 499$) of hepatitis B infection (HBV) patients, Wei et al. (13) constructed and contrasted AI draws near and the FIB-4 scoring. The HBV dataset ($n = 86$) was utilized to test the models' presentation. To test the relevance of these models, we applied them to two separate datasets of hepatitis C (HCV) ($n = 254$ and 230). Angle helping (GB) reliably beat FIB-4 scores (p b.001) and different methodologies in the revelation information for the expectation of cutting edge HF and cirrhosis. HBV-approval information showed that the GB model had a region under col-lector working trademark bend (AUROC) that was 0.918, though the FIB-4 model had a region under beneficiary working trademark bend (AUROC) of 0.841. Two HCV datasets utilized GB-based forecast and higher shorts for both GB and FIB-4 scores were important to accomplish equivalent explicitness and awareness, yet the GB-based expectation actually performed well. A correlation of the GB-based forecast technique to FIB-4 in HBV and HCV partners involving different end values for various etiological groupings showed non-stop upgrades comparative with FIB-4. LiveBoost, an easy to understand web stage, empowers our expectation models to be utilized in clinical examination and applications with practically no limitations.

Patients with ongoing hepatitis C were isolated into two gatherings in light of their METAVIR scores: those with gentle to direct fibrosis (F0–F2), and those with cut-ting edge fibrosis (F3–F4). Progressed fibrosis hazard expectation models in view of choice trees, hereditary calculations, molecule swarm advancement, and multistraight relapse calculations have been made. The proposed models were assessed utilizing collector working trademark bend examination. There were genuinely huge relationships between's cutting edge fibrosis and age, platelet count, AST, and egg whites. With an AUROC of 0.73 to 0.76 and a precision of 66.3 to 84.4 percent, the AI calculations had the option to foresee moderate fibrosis in HCC patients. Ends: Alter-native strategies, for example, AI, could be used to estimate the probability of cutting edge liver fibrosis because of constant hepatitis C disease.

As per Cai et al. (14), an outrageous learning machine was prepared to anticipate the fibrosis stage and aggravation action grade of constant hepatitis C utilizing serum lists information from patients to foster a programmed determination framework for persistent hepatitis C. For instance, the basic design and fast estimation speed of the outrageous learning machine make

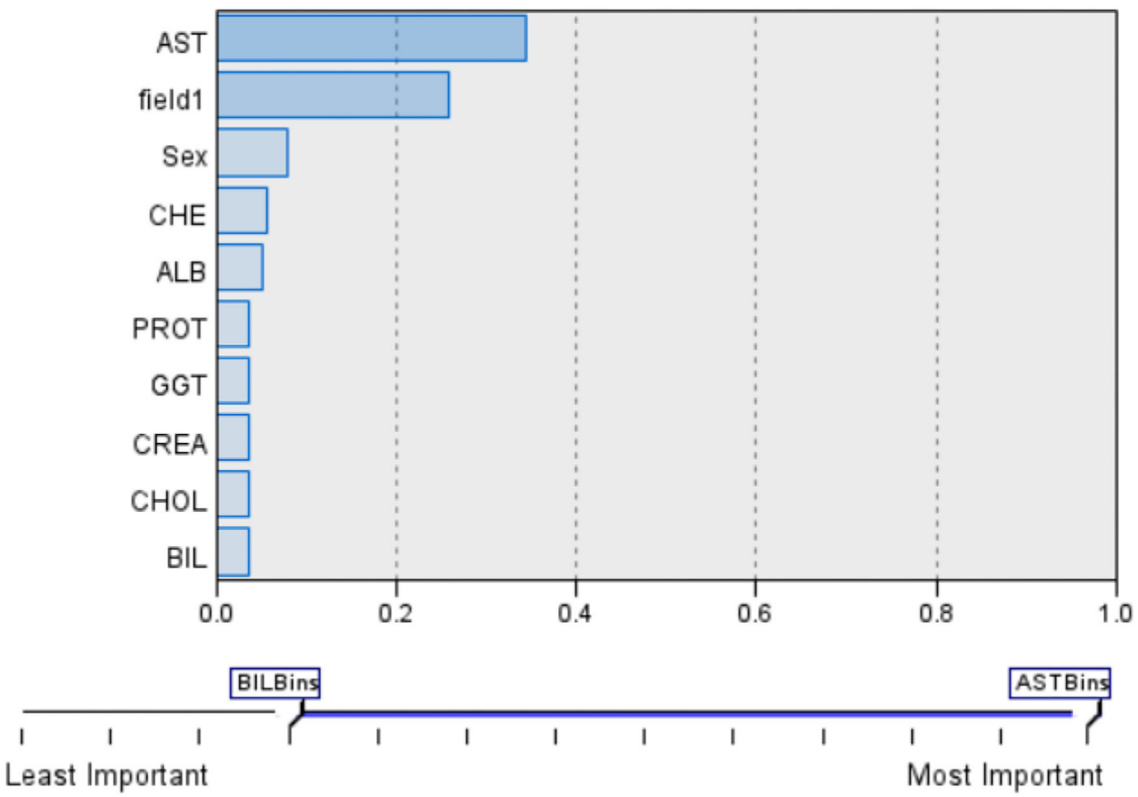


FIGURE 5 | Bayesian network Accuracy.

this independent finding framework work well. Serum markers are utilized to survey the proposed robotized determination framework for ongoing hepatitis C. For the finding of ongoing hepatitis C fibrosis stage and provocative movement grade, trial information show that the recommended method beats current baselines.

Utilizing Non-direct Iterative Partial Least Squares, Self-Organizing Map strategy, and troupes of Neuro-Fuzzy Inference System, Mehrbakhsh estimate the hepatitis infection. Likewise, we utilize choice trees to pick the most important highlights in the exploratory dataset. Utilizing a genuine world dataset, we put our procedure under serious scrutiny and contrast the outcomes with those of prior investigations. Utilizing the dataset, we observed that our strategy outflanked the Neural Network, ANFIS, K-Nearest Neighbors, and Support Vector Machines. As a shrewd learning framework for hepatitis sickness diagnostics in the medical care industry, this innovation has incredible guarantee.

Utilizing an AI approach, Ahammed et al. (15) fostered a calculation that can precisely group the periods of liver sickness in hepatitis C tainted patients. They utilized the UCI AI archive to acquire instances of liver fibrosis sickness in Egyptian patients. Manufactured minority oversampling approach has been utilized to expand the quantity of engineered patients to keep an even conveyance across all classifications. They then, at that point, utilized an assortment of element choice methods to decide the main hepatitis C viral characteristics in this dataset. Classifiers have been utilized to separate patients into bunches in view of whether their HCV cases are adjusted essential, include picked essential, or essential. In the wake of examining the information, KNN ends up as the winner, with a precision pace of 94.40%. Hepatitis C infection irresistible sickness has profited from this present review's discoveries.

NGS (cutting edge sequencing) is a usually involved strategy for delivering top caliber, profound, and proficient arrangement information. The pre-S area of the HBV genome was sequenced in 139 people, including 94 HCC patients and 45 constant HBV (CHB) patients, utilizing NGS innovation. We made two various types of information-al indexes. To begin with, we used an essential nearby arrangement search instrument (BLAST) to

plan every NGS short read and convert every arrangement into an amino corrosive by DNA codon table for the information on amino corrosive event recurrence. The info highlights are the Shannon entropy-based event frequencies of 20 major amino corrosive.

MATERIALS AND METHODS

Dataset

The dataset used in this research work was fetched from the source- Kaggle. The data set includes laboratory values of blood donors and patients suffering from Hepatitis C along with their demographic factor values such as age. The data was collected from UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/HCV$+$data.

Except Category and Sex values, all other attributes are numerical. For classification, the target attribute is Category (2): blood donors vs. Hepatitis C patients [including its progress ("just" Hepatitis C, Fibrosis, Cirrhosis)].

Methods

Data Processing

In order to transform the fetched raw data in useful and highly efficient format, certain data pre-processing techniques are employed. At this stage, different types of functions are implemented so as to find the missing values, outliers, redundant and skewed features (16).

Missing Values

Once the above pre-processed data is loaded, a function is employed to find the missing values in relation to each feature (17). The term "missing data" refers to values that aren't available, and that would be relevant if they were observed. For example, a data input error or incomplete file might cause a data set to be blank or out of sequence. Missing data is a common occurrence in real-world datasets. For analysis and modeling purposes, you must first convert data with missing data fields. This, too, might be more art than science, as is the case with many other parts of data science. Having an understanding of the data and the context in which it originates is critical. A setback isn't always a setback when you have missing numbers in your data. Although the model is lacking some information, this is a chance to do some feature engineering to help it make sense of what's been omitted. Automated detection and remediation of missing data is possible thanks to machine learning techniques and software packages.

Outliers Data

After the above pre-processing stage, a certain category of data is referred to as noisy data if it may be corrupted, distorted, or cannot be interpreted. This kind of data may have originated from improper procedures or wrong data collection. However, it can be handled by implementing the methods such as regression, clustering or binning (18–20).

Multilayer Perceptron

The feedforward artificial neural network known as a multilayer perceptron (ANN). Multi-layer perceptron (MLP) networks are commonly referred to as feedforward ANNs, but the word is also used more specifically to describe a specific type of ANN with many layers of perceptrons (with threshold activation). The input layer transmits the signal to be processed. Prediction and categorization are under the purview of the output layer. There may be any number of hidden layers between the input and output levels in the MLP's computational engine (11, 14, 21) as seen in **Figure 1**.

For the back propagation process to work, it needs to traverse two different layers of a network in both directions. Pattern or input vector are applied to the input layer, and this effect propagates across multiple layers and generates an output vector in the forward pass The weights of the synapses in the network remain constant during this operation. The weights vary during the backward pass because of the mistake correction rule. Comparing the output signal's present state to the desired state is done (13, 22) as seen in **Figure 2**.

Using MLPs, it is possible to approximate any continuous function, including those that cannot be separated linearly. Pattern classification, recognition, prediction, and approximation are the most common uses of MLP as seen in **Figure 3**.

Bayesian Network

As an alternative, developing a model that preserves known conditional dependence between random variables and conditional independence in all other situations. In a graph with directed edges, Bayesian networks represent the probabilistic graphic representation of a system's known conditional dependence. The Bayesian network seen in **Figure 6** depicts the structure of the problem domain. In order to identify hepatitis illness, the network models a variety of factors, including certain symptoms and a small number of disorder nodes (10, 15, 23–33).

Hepatitis C is more likely in those who have jaundice, and vice versa, as shown by the arcs between the two nodes: jaundice raises the likelihood of Hepatitis C, and vice versa. The model's layout serves as an illustration of the causal links between the many elements involved in the diagnostic process. **Figures 4**, **5** shows the Bayesian network for the problem under study in this paper and its accuracy.

In creating the framework, we drew inspiration from medical fiction, conversations with herpetologists, and the model's numerical elements, including as Hepatitis patients' medical records are mined for prior and conditional probability distributions. **Figure 6** shows the Bayesian network Classification for Category.

Quest

Classification systems and prediction algorithms can be built using decision tree approach, which is commonly used in data mining. Branch-like segments of a population form an inverted tree with a root node, inner nodes, and leaf nodes. Rather than relying on a complicated parametric framework, the method uses a non-parametric approach to deal with large and complex datasets. When the sample size is large enough, training and validation datasets can be segregated from one other. QUEST evaluates a node's predictor variables using a set of criteria based on significance tests. Each predictor at a node may only need to be

Results for output field Category

Comparing $B-Category with Category

Correct	581	94.47%
Wrong	34	5.53%
Total	615	

Coincidence Matrix for $B-Category (rows show actuals)

	0=Blood Donor	0s=suspect Blood Donor	1=Hepatitis	2=Fibrosis	3=Cirrhosis	$null$
0=Blood Donor	524	0	1	0	1	7
0s=suspect Blood Donor	0	7	0	0	0	0
1=Hepatitis	2	0	18	0	0	4
2=Fibrosis	1	0	3	8	0	9
3=Cirrhosis	0	0	0	0	24	6

Performance Evaluation

0=Blood Donor	0.137
0s=suspect Blood Donor	4.476
1=Hepatitis	3.043
2=Fibrosis	3.377
3=Cirrhosis	2.98

Confidence Values Report for $BP-Category

Range	0.484 - 1.0
Mean Correct	0.991
Mean Incorrect	0.742
Always Correct Above	0.979 (94.06% of cases)
Always Incorrect Below	0.484 (0% of cases)
98.64% Accuracy Above	0.0
2.0 Fold Correct Above	0.768 (99.48% of cases)

FIGURE 6 | Bayesian network Classification for Category.

tested once for selection reasons. This method does not analyze splits as thoroughly as either C&RT or CHAID does, nor does it study category combinations as thoroughly as either C&RT or CHAID does as seen in **Figure 7**.

This helps to speed things up. The target categories are divided into groups by doing a quadratic discriminate analysis using the specified predictor. When determining the ideal split, this approach is faster than C&RT (Complete and Recursive Search). Quest Diagnostics, a major clinical laboratory test supplier in the United States, provided the data for this study. Also **Figure 8** depicts the Quest Classification for Category.

Ordered HCV antibody immunoassay testing and RNA diagnostic tests for de-identified individuals. When available, data on patient gender and age were incorporated into the study. It was concluded that the data obtained from Quest Diagnostics could not be used to identify any particular person. In the event of a positive antibody test result, all specimens will be sent for HCV RNA quantitative testing.

Proposed Ensemble Learning Model

Combining predictions from multiple models to improve predictive performance is called ensemble learning. As many ensembles as you like may be created for any predictive modeling challenge, but there are only three methods that dominate the field of ensemble learning. Much to the dismay of many, each algorithm has given rise to an infinite number of sub-algorithms. It's important to understand the three main classes of ensemble learning methods before moving on to the math and programming, so that we can understand the underlying concepts.

Typical Ensemble Learning Techniques

Although there are practically infinite ways to accomplish this, there are probably three kinds of ensemble learning approaches that are most frequently studied and employed in practice. Their appeal stems from their ease of use and ability to solve a wide range of predictive modeling challenges. They are:

- Bagging.
- Stacking.
- Boosting.

Each strategy has an algorithm that specifies it, but the success of each approach has crucial to realize that, while these three methods are widely discussed and used, they don't define the scope of ensemble learning on their own.

Bagging Ensemble Learning

It is an ensemble learning method that uses bootstrap aggregation or bagging. Bootstrap and aggregation are the two primary components of Bagging, as their names suggest. An unpruned

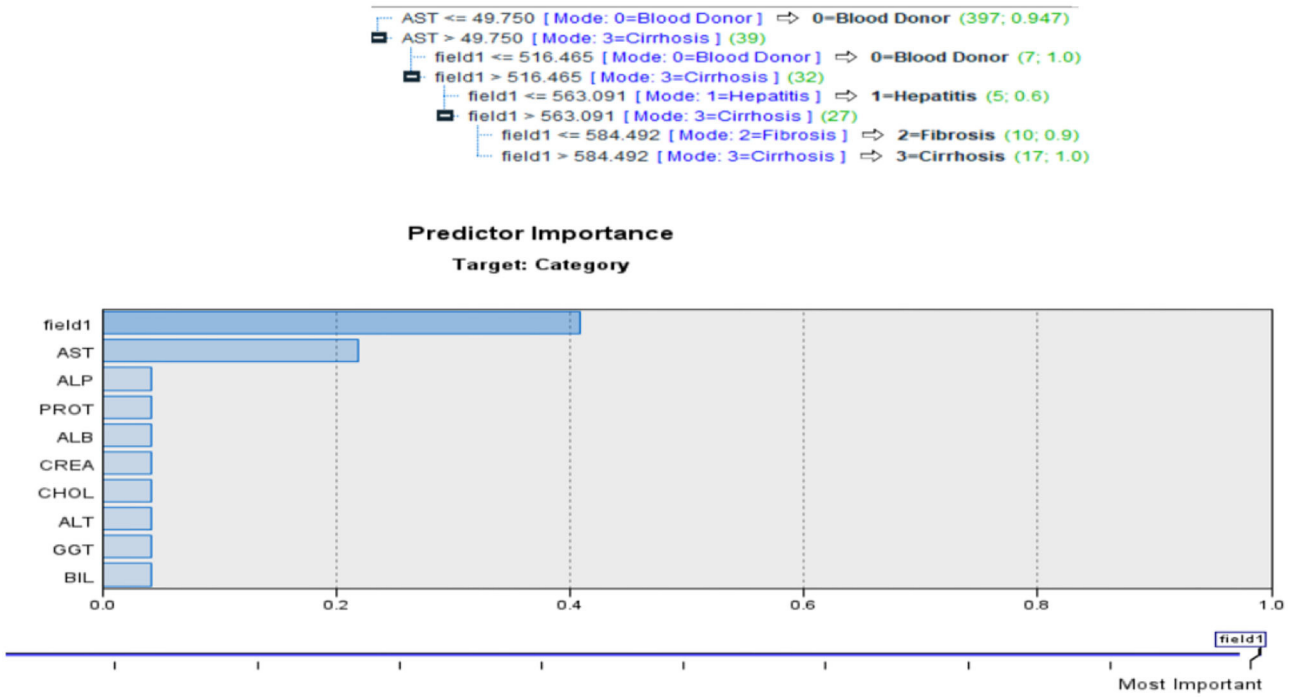


FIGURE 7 | QUEST for current problem.

decision tree is almost commonly used to train each model on a separate sample of the same training dataset in this manner. A basic statistic, such as a vote or average, is used to combine the forecasts of the ensemble members. When training the classifiers, we utilize a weak classifier whose decision boundaries vary significantly in response to even small perturbations in our training data. This ensures that the ensemble has a diverse set of classifiers.

When training ensemble members, the preparation of each dataset sample is critical. The dataset is randomly sampled for each model. In this dataset, the examples (rows) are chosen at random, but replacement is performed on them. The bootstrap distribution is used in bagging to generate alternative base learners. To put it another way, data subsets are collected via bootstrap sampling for the purpose of training the foundational learners.

It's called a bootstrap sample. In statistics, it's a method for assessing the statistical significance of a small sample of data. Rather than estimating directly from the dataset, it is possible to obtain a more accurate overall estimate of the desired quantity by doing many bootstrap sampling. Training datasets for many independent predictive models can be built in the same way and used to make predictions. Rather to fitting a single model directly to the training dataset, averaging the predictions from multiple models is often more accurate. Bagging's most important characteristics can be summarized as follows:

- Samples from the training dataset that have been bootstrapped.
- On each sample, unpruned decision trees fit.
- Simple voting or prediction averaging.

In summary, bagging makes a contribution by changing the training data used to fit each ensemble member, resulting in skilled but distinct models.

Bagging Ensemble

It's a broad strategy that can be readily expanded. More changes to the training dataset, for example, could be made, the algorithm that fits the training data could be altered, and the mechanism that combines predictions might be changed. This method is used in a number of popular ensemble algorithms, including:

- Bagged Decision Trees (canonical bagging)
- Random Forest
- Additional Trees

Next, let's look into stacking in more detail.

Stacking Ensemble Learning

Stacked To find a diverse group of members via generalization or stacking, one can vary the model types that are fit on training data and combine these models' predictions with one another. Stacking is a common technique for teaching a student how to mix up a large group of students. In education, the term "first-level learner" refers to an individual student, whereas "second-level learner" refers to the combined group. Level-0 models refer to the members of an ensemble, while level-1 models describe the

Results for output field Category

Comparing $R-Category with Category

Correct	582	94.63%
Wrong	33	5.37%
Total	615	

Coincidence Matrix for $R-Category (rows show actuals)

	0=Blood Donor	1=Hepatitis	2=Fibrosis	3=Cirrhosis
0=Blood Donor	533	0	0	0
0s=suspect Blood Donor	4	3	0	0
1=Hepatitis	13	10	1	0
2=Fibrosis	6	0	14	1
3=Cirrhosis	5	0	0	25

Performance Evaluation

0=Blood Donor	0.092
1=Hepatitis	2.981
2=Fibrosis	3.308
3=Cirrhosis	2.981

Confidence Values Report for $RC-Category

Range	0.4 - 0.938
Mean Correct	0.912
Mean Incorrect	0.877
Always Correct Above	0.938 (0% of cases)
Always Incorrect Below	0.4 (0% of cases)
94.63% Accuracy Above	0.0
2.0 Fold Correct Above	0.938 (0% of cases)

FIGURE 8 | Quest Classification for Category.

model used to integrate the forecasts. The following are a few of the most important aspects of stacking:

- The training dataset has not been modified.
- For each ensemble member, different machine learning techniques are used.
- Using a machine learning model, we can figure out how to combine predictions in the most effective way.

The several machine learning models that make up the ensemble add variety. Using a number of models that are taught or developed in a variety of ways ensures that they make a wider range of assumptions and, as a result, have less related predictions errors (34, 35). In an ensemble method known as "boosting," the training data is manipulated in an attempt to draw attention to examples that previous models that were fitted to the training dataset mistakenly detected. In boosting, the training dataset for each new classifier is progressively narrowed to examples that prior classifiers had incorrectly classified. The idea of correcting prediction errors is an important characteristic of boosting ensembles. To make sure that the first model's predictions are as accurate as possible, subsequent models are fitted and added to the ensemble in a logical order. Weak learners, or decision trees that only make one or a few judgments, are commonly used to do this.

Data can be weighed to signify how much attention an algorithm should give the model while it is being learnt. The following are a few of the most important aspects of boosting:

- Bias training data in favor of difficult-to-predict examples.
- Using a weighted average of models, combine forecasts.

There have been a few different approaches to making it possible to turn a large number of weak students into a small number of strong ones. The Adaptive Boosting (AdaBoost) method proved boosting to be an effective ensemble strategy for this paper. An algorithm known as "boosting" can predict more accurately.

RESULTS AND DISCUSSION

In this research paper, following machine learning algorithms are being used including:

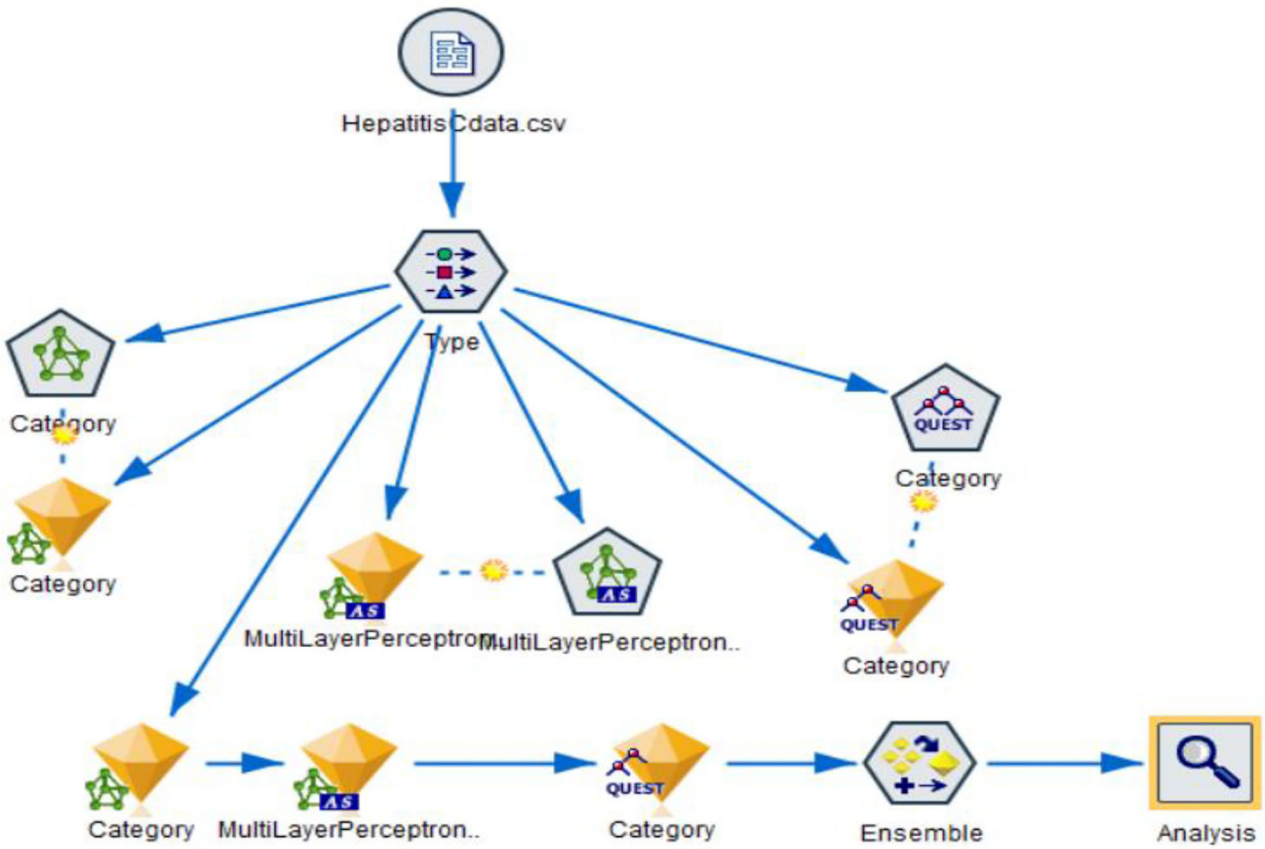


FIGURE 9 | Proposed Ensemble learning model.

- MLP
- Bayesian Network
- QUEST

The **Figures 9**, **10** depicts the working of the proposed Ensemble Learning model. Ensemble node mixes three model nuggets to obtain more accurate predictions than can be derived from any of the individual models (MLP, Bayesian Network & QUEST) (MLP, Bayesian Network & QUEST). By merging predictions from different models, limitations in MLP, Bayesian Network & QUEST models have been eliminated, resulting in a higher overall accuracy. MLP, Bayesian Network & QUEST Models integrated in this manner often perform at least as well as the best of the MLP, Bayesian Network & QUEST models and often better.

Table 1 represents the level of accuracy at 95.59%. This value is more efficient to individual machine learning algorithms as shown in **Table 1**. This is also represented in **Figure 11**.

Then results of this proposed model indicate that the proposed ensemble model can accurately predict the advanced fibrosis stage in chronic HCV patients. The proposed model offers us 95.9% correct results. We can improve the model's efficiency in the future by adding more patient data. In addition, more liver ailments could be investigated in the future. Figures in the paper appeared in sequence.

TABLE 1 | Accuracy comparison.

S. No.	Name of Algorithm	Accuracy (%)
1	MLP	94.10
2	Bayesian Network	94.47
3	QUEST	94.63
4	Proposed Ensemble Model	95.59

CONCLUSION

In this research paper, different machine learning algorithms are being applied for predicting advanced liver fibrosis in Chronic Hepatitis C patients. It is observed that individual model is capable of providing the accuracy up to 94.67%. Then the ensemble model including Bayesian network, MLP and QUEST decision trees has been developed.

In designing healthcare systems, innovative techniques that carefully balance public health initiatives with limited resources should be taken into consideration. In order to help individuals who don't have health insurance or are underinsured, there should be an

Results for output field Category

Comparing $XS-Category with Category

Correct	563	95.59%
Wrong	26	4.41%
Total	589	

FIGURE 10 | Accuracy level of Proposed Ensemble learning model.

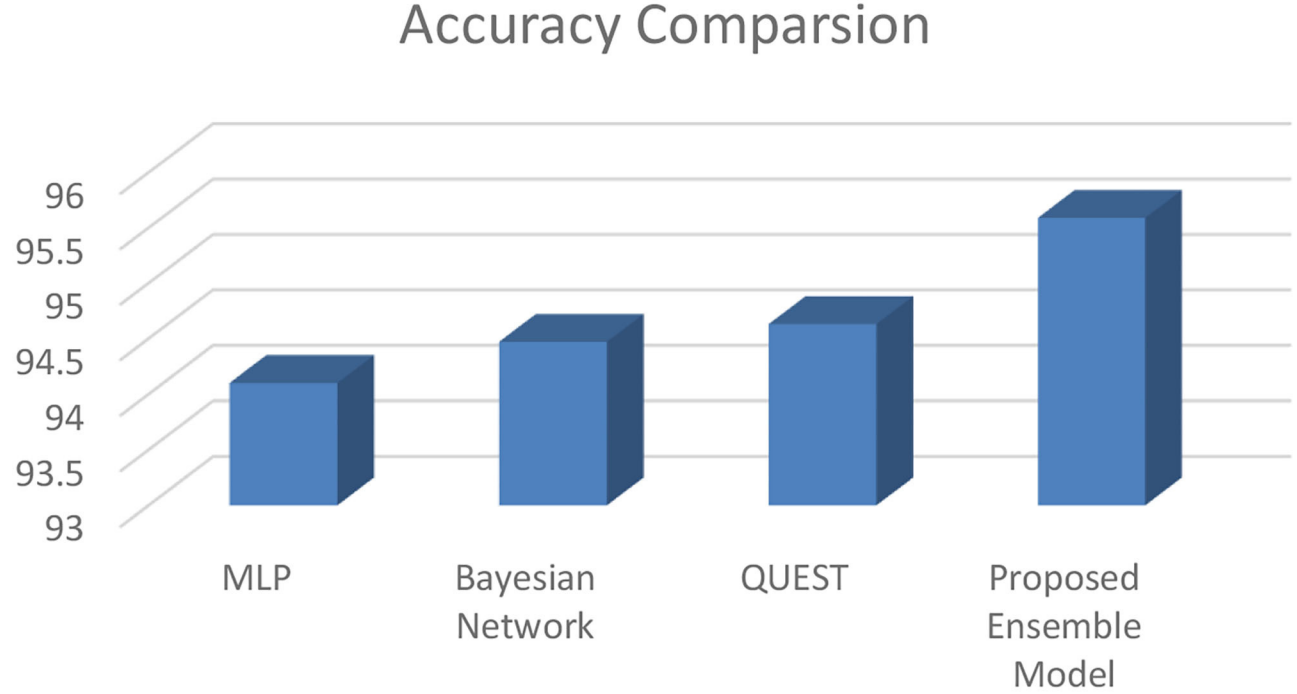


FIGURE 11 | Accuracy comparison.

increase in access to healthcare, community outreach, and the growth of telemedicine, including safe laboratory testing.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://archive.ics.uci.edu/ml/datasets/HCV+data.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

The authors are grateful for the financial support provided by Dar Al-Hekma University, Jeddah, Saudi Arabia.

REFERENCES

1. Ibrahim H, El Kerdawy AM, Abdo A, SharafEldin A. Similarity-based machine learning framework for predicting safety signals of adverse drug–drug interactions. *Informatics Med.* (2021) 26:100699. doi: 10.1016/j.imu.2021.100699
2. Nakayama JY, Ho J, Cartwright E, Simpson R, Hertzberg VS. Predictors of progression through the cascade of care to a cure for hepatitis C patients using decision trees and random forests. *Comput Biol Med.* (2021) 134:104461. doi: 10.1016/j.compbiomed.2021.104461
3. Stepanova M, Younossi I, Racila A, Younossi ZM. Prediction of health utility scores in patients with chronic hepatitis C using the Chronic Liver

Disease Questionnaire-Hepatitis C Version (CLDQ-HCV). *Value Heal.* (2018) 21:612–21. doi: 10.1016/j.jval.2017.10.005
4. Ma L, Yang Y, Ge X, Wan Y, Sang X. Prediction of disease progression of chronic hepatitis C based on XGBoost algorithm. In: *2020 International Conference on Robots & Intelligent System (ICRIS).* (2020). pp. 598–601.
5. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Heal.* (2021) 20:100178. doi: 10.1016/j.smhl.2020.100178
6. Abd El-Salam SM, Ezz MM, Hashem S, Elakel W, Salama R, ElMakhzangy H, et al. Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Informatics Med.* (2019) 17:100267. doi: 10.1016/j.imu.2019.100267
7. Wong GLH, Hui VWK, Tan Q, Xu J, Lee HW, Yip TCF, et al. Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis. *JHEP Rep.* (2022) 4:100441. doi: 10.1016/j.jhepr.2022.100441
8. Sun H, Zhang A, Yan G, Piao C, Li W, Sun C, et al. Metabolomic analysis of key regulatory metabolites in hepatitis c virus-infected tree shrews. *Mol Cell Proteomics.* (2013) 12:710–19. doi: 10.1074/mcp.M112.019141
9. Akbilgic O, Obi Y, Potukuchi PK, Karabayir I, Nguyen DV, Soohoo M, et al. Machine learning to identify dialysis patients at high death risk. *Kidney Int Rep.* (4:1219–29. doi: 10.1016/j.ekir.2019.06.009
10. ElHefnawi M, Abdalla M, Ahmed S, Elakel W, Esmat G, Elraziky M, et al. Accurate prediction of response to Interferon-based therapy in Egyptian patients with Chronic Hepatitis C using machine-learning approaches. In: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.* Istanbul (2012). pp. 771–8.
11. Wei Y, Li J, Qing J, Huang M, Wu M, Gao F, et al. Discovery of novel hepatitis C virus NS5B polymerase inhibitors by combining random forest, multiple e-pharmacophore modeling and docking. *PLoS ONE.* (2016) 11:e0148181. doi: 10.1371/journal.pone.0148181
12. Singh A, Mehta JC, Anand D, Nath P, Pandey B, Khamparia A. An intelligent hybrid approach for hepatitis disease diagnosis: combining enhanced k-means clustering and improved ensemble learning. *Expert Syst.* (2021) 38:e12526. doi: 10.1111/exsy.12526
13. Wei R, Wang J, Wang X, Xie G, Wang Y, Zhang H, et al. Clinical prediction of HBV and HCV related hepatic fibrosis using machine learning. *EBioMedicine.* (2018) 35:124–32. doi: 10.1016/j.ebiom.2018.07.041
14. Cai J, Chen T, Qiu X. Fibrosis and inflammatory activity analysis of chronic hepatitis c based on extreme learning machine. In: *2018 9th International Conference on Information Technology in Medicine and Education (ITME).* Hangzhou (2018). pp. 177–81.
15. Ahammed K, Satu MS, Khan MI, Whaiduzzaman M. Predicting infectious state of hepatitis c virus affected patient's applying machine learning methods. In: *2020 IEEE Region 10 Symposium (TENSYMP).* Dhaka (2020). pp. 1371–4.
16. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, et al. Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. *EBioMedicine.* (2019) 50:156–65. doi: 10.1016/j.ebiom.2019.10.057
17. Huang H, Shiffman ML, Friedman S, Venkatesh R, Bzowej N, Abar OT, et al. A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C. *Hepatology.* (2007) 46:297–306. doi: 10.1002/hep.21695
18. Zidan AM, Saad EA, Ibrahim NE, Hashem MH, Mahmoud A, Hemeida AA. Host pharmacogenetic factors that may affect liver neoplasm incidence upon using direct-acting antivirals for treating hepatitis C infection. *Heliyon.* (2021) 7:e06908. doi: 10.1016/j.heliyon.2021.e06908
19. Ghazal TM, Anam M, Hasan MK, Hussain M, Farooq MS, Ali HM, et al. Hep-pred: hepatitis C staging prediction using fine gaussian SVM. *Comput Mater Contin.* (2021) 69:191–203. doi: 10.32604/cmc.2021.015436
20. Chen L, Lu J, Huang T, Yin J, Wei L, Cai YD. Finding candidate drugs for hepatitis C based on chemical-chemical and chemical-protein interactions. *PLoS ONE.* (2014) 9:e107767. doi: 10.1371/journal.pone.0107767
21. Abd-Elsalam SM, Ezz MM, Gamalel-Din S, Esmat G, Salama A, ElHefnawi M. Early diagnosis of esophageal varices using Boosted-Naïve Bayes Tree: a multicenter cross-sectional study on chronic hepatitis C patients. *Informatics Med.* (2020) 20:100421. doi: 10.1016/j.imu.2020.100421
22. Hashem S, Esmat G, Elakel W, Habashy S, Raouf SA, Elhefnawi M, et al. Comparison of machine learning approaches for prediction of advanced liver fibrosis in chronic hepatitis C patients. *IEEE/ACM Trans Comput Biol Bioinform.* (2018) 15:861–8. doi: 10.1109/TCBB.2017.2690848
23. Safdari R, Deghatipour A, Gholamzadeh M, Maghooli K. Applying data mining techniques to classify patients with suspected hepatitis C virus infection. *Intell Med.* (2022) 19:124–132. doi: 10.1016/j.imed.2021.12.003
24. Chávez-Tapia NC, Ridruejo E, Alves de Mattos A, Bessone F, Daruich J, Sánchez-Ávila JF, et al. An update on the management of hepatitis C: guidelines for protease inhibitor-based triple therapy from the Latin American Association for the Study of the Liver. *Ann Hepatol.* (2013) 12:S3–5. doi: 10.1016/S1665-2681(19)31404-8
25. Chicco D, Jurman G. An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis. *IEEE Access.* (2021) 9:24485–98. doi: 10.1109/ACCESS.2021.3057196
26. Onyema EM, Piyush KS, Surjeet D, Mayuri NM, Mohammed Z, Basant T. Enhancement of patient facial recognition through deep learning algorithm: ConvNet. *Hindawi J Healthc Eng.* (2021) 2021:5196000. doi: 10.1155/2021/5196000
27. Onyema EM, Elhaj MAE, Bashir SG, Abdullahi I, Hauwa AA, Hayatu AS. Evaluation of the performance of K-nearest neighbor algorithm in determining student learning styles. *Int J Innov Sci Eng Techn.* (2020) 7:91–102. doi: 10.1109/TENSYMP50017.2020.9230464
28. Iwendi CC, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357
29. Alazab M, Lakshmanna K, Reddy T, Pham QV, Maddikunta PKR. Multi-objective cluster head selection using fitness averaged rider optimization algorithm for IoT networks in smart cities. *Sustain Energy Technol Assess.* (2021) 43:100973. doi: 10.1016/j.seta.2020.100973
30. Jemmali M, Denden M, Boulila W, Jhaveri RH, Srivastava G, Gadekallu TR. A novel model based on window-pass preferences for data-emergency-aware scheduling in computer networks. In: *IEEE Transactions on Industrial Informatics.* IEEE (2022).
31. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran, et al. A metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw Pract Exp.* (2021) 51:2558–71. doi: 10.1002/spe.2797
32. Onyema EM, Tariq AA, Ghouali S, Manish S, Manish M, Guellil MS, et al. Empirical analysis of apnea syndrome using an artificial intelligence-based granger panel model approach. *Comput Intellig Neurosci.* (2022) 2022:7969389. doi: 10.1155/2022/7969389
33. Onyema EM, Khalaf OI, Romero CAT, Tayeb S, Ghouali S, Abdulsahib GM, et al. A classification algorithm-based hybrid diabetes prediction model. *Front Public Health.* (2022) 10:829519.
34. Dhaou BI, Ebrahimi M, Ammar BM, Bouattour G, Kanoun O. Edge devices for internet of medical things: technologies, techniques, and implementation. *Electronics.* (2021) 10:2104. doi: 10.3390/electronics10172104
35. Nguyen TG, Ali M, Dhaou IB, Rahmani AM, Westerlund T, Liljeberg P, et al. IoT-based continuous glucose monitoring system: a feasibility study. *Proc Comput Science.* (2017) 109:327–34. doi: 10.1016/j.procs.2017.05.359

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Edeh, Dalal, Dhaou, Agubosim, Umoke, Richard-Nnabu and Dahiya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fpubh.2022.897459

Testing the Stochastic Convergence of Ecological Indicators in BRICS in the Perspective of Public Health

*Jinshun Wu**

Department of Economics, School of Economics and Management, East China Jiaotong University, Nanchang, China

This paper assesses the stochastic convergence of relative per capita ecological footprints within BRICS countries over the period 1961–2017 in the field of Public Health. Our initiatives have targeted ecological Indicator and health behaviors. Using the local Whittle estimator and some of its variants we assess whether relative per capita ecological footprints are long memory processes which, although highly persistent, may revert to their mean/trend in the long run thereby indicating evidence of stochastic convergence, or divergent processes in nutshell. Furthermore, we test whether (possibly) slow convergence or the complete lack of it may be the result of structural changes to the deterministics of each of the relative per capita footprint series by means of the tests of Berkes et al. (1) and Mayoral (2). For the ease of comparison, this paper assesses the stochastic convergence of relative per capita ecological capacities for BRICS as well. Our results show relatively strong evidences against stochastic convergence of ecological footprints. Furthermore, with regard to China and Russia, our results also decisively indicate that a slow or lack of convergence is the results of a structural break in the relative per capita ecological footprint series. However, our empirical researches support stochastic convergence of relative per capita ecological capacities for BRICS. In addition, we can conclude whether the per capita ecological footprints converge or not is dependent on the level of economic development, and the stochastic convergence occurs in those rich countries more probably, indicating that public health is becoming a more serious concern in developing countries.

Keywords: BRICS, ecological indicator, fractionally integrated processes, stochastic convergence, structural break

JEL Classification: C12, C14, Q50, Q54

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
B. V. Santhosh Krishna, New Horizon College of Engineering (NHCE), India
Kuruva Lakshmanna, VIT University, India

***Correspondence:**
Jinshun Wu
wujinshun6717@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *16 March 2022*
Accepted: *06 April 2022*
Published: *29 April 2022*

Citation:
Wu JS (2022) Testing the Stochastic Convergence of Ecological Indicators in BRICS in the Perspective of Public Health. Front. Public Health 10:897459. doi: 10.3389/fpubh.2022.897459

INTRODUCTION

The convergence implication envisaged by the neoclassical growth model is that if each country has the same savings rate, population growth rate and technological progress rate, the income level gap between countries will gradually disappear. Since the production of most goods and services requires the consumption of energies such as coal, natural gas and oil, energy-producing requires the exploitation of natural resources. The availability of many natural resources at low cost increases the vulnerability of countries' economic development models. Considering the close relationship between economic growth and environmental pollution, Grossman and Krueger (3) first introduced the Environmental Kuznets Curve (EKC). The EKC indicates that the degree of

environmental deterioration will increase with the increase of the per capita income level, but the environmental quality will start to increase when the per capita income level reaches a certain threshold. This hypothesis means that there is an inverted U-shaped relationship between per capita income and environmental impact indicators and economic growth. At the same time, the view that economic growth will promote environmental improvement after a certain critical point has been discussed in many studies. Many studies believe that the income level of developing countries has not reached a critical point. With an increase in incomes, environmental pollution will continue to rise. On the other hand, developed countries are more successful in reducing environmental pollution, as their higher income levels may exceed the critical point or adopt cleaner energy technologies. Brock and Taylor (4) linked the environmental pollution indicators with the EKC hypothesis. Their Catch-up hypothesis means that poor countries have more environmental pollution. Moreover, in the initial stage, the gap of environment indicators between the rich and poor countries is divergent. This difference is rooted in the difference in the initial capital stock of the two countries. However, with their incomes increasing, once the developing countries start to use the advanced technology of environmental protection, the difference in environmental quality between the two countries will be reduced and various environmental pollution indicators will converge. Although the initial stage of economic growth resulted in environmental damage, the environmental quality is bound to be improved eventually.

In the past, air pollutants were used as a single indicator to measure environmental quality, which could not fully reflect the improvement of environmental quality. Now there is a more comprehensive environmental indicator, the ecological footprint. Ress (5) first proposed the concept of ecological footprint (EF), which was then developed by 4. EF is a measure of human demand for natural resources and services, revealing the relationship between human lifestyle, consumption patterns and consumed natural capital (5), which can comprehensively evaluate the feasibility of achieving a country's sustainable development goals. The Global Ecological Footprint Network describes the ecological footprint as: the ecological footprint measures the amount of "biologically productive" land or water that enables human beings to sustain itself. The term "biologically productive areas" refers to the areas with biological productivity that are needed to maintain the survival of individuals, regions or countries, or that can absorb wastes discharged by human beings. The ecological footprint can be used to evaluate and manage resource use throughout the country, measure and assess if people's lifestyles are sustainable, and account how much waste can be absorbed by the ecosystem, also known as "appropriated carrying capacity." EF consists of 6 sub-footprints: carbon footprint, fishing grounds footprint, forest footprint, cropland footprint, built-up land footprint and grazing land footprint. These ecological footprints reflect a country's ownership of natural resources and the ecologically productive area that has been used, and can comprehensively reflect the environmental problems as well. Moreover, the ecological footprint is a resource accounting tool that measures how much renewable energies in the biosphere can be used by human activities. In addition, EF can also measure the ecological cost of goods and services that the nature provides to humans, such as land, and the maximum sustainable population in a given area. In a word, EF indicates the environmental limit and the extent to which human beings exceed the natural limit, which can assist countries to evaluate how the ecological resources are being used.

This study focuses on the BRICS countries for the following reasons:

First, compared with many other emerging countries, the BRICS countries have experienced a rapid transition from an ecological surplus to an ecological deficit. This is mainly due to the rapid economic growth of these countries over the past decade. The BRICS countries contribute 21% of global GDP and have 41% of the world's population. These countries have an average annual economic growth rate of 6.5% (6). Moreover, between 2005 and 2016, the total GDP of the BRICS countries ranged from US$ 218.7 billion to US$ 162.6 billion. The rapid economic growth has also led these countries to consume more than 40% of the global total energies, and they are also the major global CO_2 emitters (7). They are facing both environmental and natural resource stresses due to their rapid economic growth. Their ecological indicators may be long-memory and may not converge. In view of this, in order to reverse the growing trend of carbon dioxide emissions, reasonable policies are needed to ensure that while economic development is taking place, it will not cause huge damage to the ecological environment.

Secondly, despite the huge potentials of renewable energies, BRICS countries still rely excessively on fossil fuels to meet their energy needs, which poses a huge threat to the environment. At the same time, while pursuing high-speed economic development, the region has also experienced a continuous decline in ecological sustainability. In addition, the research on this area can help the policy makers to prevent the environmental deterioration with practical and reasonable policies.

LITERATURE REVIEW

The literature on the deterioration of ecological footprints mainly focuses on developed countries and international organizations mainly composed of developed countries. As far as we know, the first paper on convergence of ecological footprints is Ulucak and Lin (8). They analyzed the random behavior of ecological footprints and each sub-component using Fourier unit root test *via* the dataset of ecological footprints of each state in the United States from 1961 to 2012. The empirical results show that the ecological footprints of states in US are non-stationary, and not mean-reverting processes. Therefore, an impact on these sequences will result in a permanent deviation from the long-term equilibrium paths. Since then, there have been a lot of relevant researches mainly exploring developed countries' ecological indicators. For example, Ulucak and Apergis (9) studied the convergence of the per capita ecological footprints using data samples from 20 EU countries during the period 1961–2013. The results show that there are three convergence clubs. Bilgili and Ulucak (10) studied the convergence of per capita

ecological footprints among G20 countries during 1961–2014 by using bootstrap panel KPSS unit root test and club convergence test considering structural breakpoints. It is concluded that the per capita ecological footprints of these countries will reach random convergence and deterministic convergence, and two convergence clubs are identified. Solarin (11) uses LM and RALS-LM unit root methods to test the stochastic convergence of ecological footprints of 27 OECD countries from 1961 to 2013. They found that the ecological footprints and carbon footprints of 25 countries are stationary. They also performed β convergence and σ convergence test in their research. Among the 25 countries with random convergence, the ecological footprints of 13 countries are β convergent; among them, 15 countries have β convergent carbon footprints. In terms of σ convergence, the ecological footprints and carbon footprints of all countries selected meet the convergence criteria. Studies show that, although not all countries meet the convergence conditions, those countries in which convergence happened already can jointly formulate environmental protection policies to improve convergence among more countries. Yilanci et al. (12) uses the annual data of ecological footprints and each sub-component of 25 OECD countries from 1961 to 2013, and uses various stability test methods to prove that the other footprint sequences except the fishing ground footprint are stationary. Yildirim et al. (13) proposed the Fourier cross-section panel KSS unit root test to study the convergence of ecological footprints among 16 EU membership countries. They further use the rolling window method to consider the time-varying stationarity of the ecological footprint sequences. Through their study on the convergence of the ecological footprints and each sub-component, they prove that the convergence or divergence of the ecological footprints has different states in different periods.

In recent years, there have been a large number of researches on the convergence of ecological footprints between countries with different levels of development, and the relationship between the convergence of ecological footprints and the degree of economic development is discussed. Solarin and Bello (14) studied the convergence of the ecological footprints of 128 countries with different levels of development during the period 1961–2013. The results show that the ecological footprint series of 96 countries (81%) are non-stationary and there is no mean reversion (non-convergence). This means that various environmental protection policies have long-term and permanent effects in many countries. Bilgili et al. (15) selected 15 countries in Asia, Africa, America and Europe to test the stationarity of the ecological footprints, and asserted that the ecological footprints of the Asian group are non-stationary and diverge. Using the club convergence method proposed by Phillips and Sul (16) and Solarin et al. (17) showed that 92 countries' ecological footprints and each sub-component have club convergence. Ozcan et al. (18) used panel KSS and Fourier panel KSS to study the stability and convergence of ecological footprints in 113 countries with different levels of development depending on their national development levels. The ecological footprints of all high-income countries, some low-income countries and upper-middle-income countries are stable, while the ecological footprints of lower-middle-income countries are not stationary. Erdogan and Okumus (19) used the annual sample data from 1961 to 2016 to test the random convergence of ecological footprints and club convergence of different income groups. The FPKPSS panel statistics of the countries in the high, middle and low income groups indicate that the per capita ecological footprints are non-stationary. At the same time, the FKPSS method was also used to test the stability of ecological footprints at the national level. On this occasion, 6 out of 26 high-income countries and 8 out of 38 middle-income countries have stationary ecological footprints. In 8 out of 25 low-income countries ecological footprints are stationary. At the same time, the club convergence method was used to test the convergence of the ecological footprints. It is concluded that there are convergence clubs in different income groups. Işik et al. (20) employed the threshold autoregressive panel unit root test to study the convergence of ecological footprints of NAFTA countries during 1961–2016. The study suggested that EF converges at the latter stage of the threshold, and the second stage accounts for 48.8% of the total sample. The ecological footprints of the first stage are divergent. These conclusions highlight the need for NAFTA countries to formulate common environmental policies to mitigate environmental degradation.

With the rapid economic development in developing countries, the academic community has become more and more interested in the convergence of the ecological footprints of international organizations composed of developing counties in recent years. Yilanci and Pata (21) used the panel data of the five ASEAN countries' per capita ecological footprints from 1961 to 2016, and used two-mechanism threshold autoregressive panel unit root test to study the convergence of ecological footprints. Researches argue that the ecological footprints in the second mechanism are divergent; in the first mechanism they are absolutely convergent. The second mechanism accounts for 80% of the sample, so the results strongly support that the per capita ecological footprints of the five ASEAN countries are absolutely convergent, and they believe that common policies should be implemented to slow down the environmental deterioration. Tillaguango et al. (22) studied the convergence of the ecological footprints of Latin American countries from 1990 to 2016 and found that there exits a per capita ecological footprint club. Using the method of logistic regression, the paper explored the factors that affect the convergence of the ecological footprint clubs and found that economic complexity, shadow economy and abundance of natural resources can significantly affect the formation of the clubs. Their research indicated that there are three convergence clubs in Latin America, and the two clubs are divergent.

As can be seen from the literature review above, a large number of techniques have been applied to test the stochastic convergence and divergence of ecological and environmental indicators (8, 10, 13, 18, 20, 23, 24), but most of these studies only used the time series econometric method to analyze the existence of its stochastic convergence characteristics. From a random point of view, if the per capita ecological footprint sequences tend to converge as time lapses, the impacts of external shocks on the sequence should be only temporary, and the data are stationary. In the presence of a unit root, the external

impacts on the sequence will be permanent, indicating that these sequences are divergent. From the literature review, it can also be clearly seen that the evidence of convergence of ecological and environmental indicators is very complicated. One possible explanation is that the existing literature ignores the possibility that ecological footprints are long memory processes, i.e., they are mean- (or trend-) reverting, but it will take a long time for these processes to arrive at their means or trends. There are two reasons to doubt this view: First, a country's ecological footprints are determinated by a lot of fundamental factors, such as economic size, economic structure and technological level. All these factors may only change slowly as time goes by, so environmental indicators such as relative ecological footprints between countries may also change slowly. Secondly, previous studies have found that GNP and per capita GNP are long memory processes (2). The long memory process means that there exists significant interdependent relationship between two observations over a long time, so the effects of shocks tend to decay slowly, although they are still mean-reverting in nature. If environmental indicators such as ecological footprints are indeed long memory processes, then previous studies might provide misleading results. For example, if the null hypothesis of a unit root has not been rejected in previous studies, this may not mean a lack of convergence, but may reflect that the environmental indicator has the difference order of fractionally integrated processes, which might not be integer. Therefore, this means that the convergence process is taking place, albeit vary slowly.

The analysis in this paper focuses on conditionally stochastic convergence, which does not require every country (region) to converge to the same steady-state level just as β and σ convergence do. When the per capita ecological footprint gap between countries (regions) follows a mean stationary process, i.e., the impact of the relative per capita ecological footprints only brings temporary deviation from their convergent trends. This paper uses the stochastic convergence test approach to determine whether it can be ensured that ecological indicators converge permanently in a country. If such a necessary mechanism does not exist, deviations from trends are permanent. Thus, the concept of convergence is more important to the study of ecological footprints because it can test the durability of impact effects. In this regard, few researchers use unit root test to study whether the impacts on ecological indicators are temporary or permanent. The lack of convergence researches makes it difficult for a country to take joint actions to hinder environmental deterioration. The stochastic behavior and dynamic changes of ecological indicators will help to formulate sustainable development policies. If ecological indicators are stationary, their impacts exerted by external shocks are transient, and these sequences will converge stochastically to their means once these impacts disappear. On the contrary, if the impacts are permanent, the sequences will diverge from their means. If the evolution trend of variables is known, the convergence and divergence of environmental indicators make it easier for policymakers to apply more effective environmental protection policies. In this paper, the fractionally integrated autoregressive moving average model (ARFIMA) is used to estimate the fractionally integrated parameters of the ecological indicators using the 1961–2017 data set. The ARFIMA model is a generalized version of ARIMA and an improvement on the I (0) |I (1) dichotomy. It can estimate the integration order parameters, which can be of any real number in the data generated processes. Because the unit root test can only judge whether the ecological indicators are convergent I (0) processes or divergent I (1) processes, and cannot test the convergence speed, the fractionally integrated model may be more appropriate.

In this paper, for exploring whether the relative ecological indicators per capita contain unit roots (the null hypothesis is a non-convergent process). We leverage the local Whittle (LW) estimation method (25) and its modified versions, e.g., exact local Whittle (ELW) and two-stage detrended (breakpoint) exact local (ELW) estimation method (26, 27) to estimate the parameters of fractionally integrated model. In contrast, the alternative assumption is that they are long-term memory processes that, although highly persistent, may return to their mean/trend over the long term (meaning slow convergence). The existence of structural breakpoints may lead to biased estimation of parameters [see (28, 29)]. In the case of the fractionally integrated processes, this problem is mostly applied to evaluate their expectation performance. From our point of view, there is evidence that the existence of structural breakpoints on the deterministic trend may lead to an overestimation of the order of the fractionally integrated process, so the conclusions drawn are biased in favor of supporting the long memory process [see (30, 31)]. In this study, the structural breakpoints are incorporated into the ARFIMA model by introducing dummy variables into the deterministic components Z_t of Equation (5), including mean breakpoints and trend breakpoints, respectively. Furthermore, the Mayoral (2) test is used to identify the false long memory state in the ecological indicator sequences, i.e., the long memory phenomenon is caused by the existence of breakpoints in the deterministic trend of the sequences. Finally, we use the non-parametric CUSUM statistics of Berkes et al. (1) to determine the points in time when the structural changes of the ecological indicator sequences occur (if there is a breakpoint).

MODEL

In this paper, the fractionally integrated models are used to test whether the environmental indicators, e.g., ecological footprints and ecological carrying capacities in BRICS converge. The idea of this test is that if the per capita environmental indicators are converging between countries, the logarithm of the ratio of the per capita environmental indicators in a country to the average environmental indicators in all countries should be stationary, or at least the mean- (or trend-) reverting. Therefore, referring to Strazicich and List (32), this paper defines y_{it} as the natural logarithm value of the relative environmental indicators per capita in t year of the i country, and its expression is shown in

Equation (1):

$$y_{it} = \ln\left(\frac{PCE_{it}}{J^{-1}\sum_{j=1}^{J} PCE_{it}}\right) \quad (1)$$

The PCE_{it} is the per capita environmental indicator in t year for the i country. J is the total number of countries. In order to test the null hypothesis that per capita environmental indicators diverge, we examine whether the natural logarithm y_{it} of the annual per capita relative environmental indicators in the i country contain a unit root, or whether they can be characterized by trend-stationary and mean-reverting processes.

Discrimination and Test of Convergence of Relative Environmental Indicators per Capita

When a time series has the structural change in its deterministic trend, the standard unit root test method often has the problem of significant level distortion; Moreover, when the unit root of the sequence is close to 1, the performance of standard unit root test is affected substantially. In other words, the unit root test is often unable to decide whether a sequence is highly persistent or infinitely persistent. The test method based on fractionally integrated processes provides a general framework, which can test whether an ecological and environmental indicator is an I (1) process with infinite persistence or with long persistence (which may be not covariance-stationary), but characterized by mean reversion in the long term, indicating that the environmental indicator is an $I(d)$ process with the integration order bounded the domain that $0 < d < 1$. The fractionally integrated expression of a environmental indicator sequence containing a time trend, such as the ARIMA (0, d, 0) model, is shown by Equation (2):

$$y_{it} = c_{i0} + \gamma_i t + x_{it} \quad (2)$$

$$(1-L)^{d_i} x_{it} = u_{it} \quad (3)$$

where u_{it} is a white noise process with zero mean value, d_i is the integration order of a single sequence and a deterministic trend function (there may be breakpoints). According to various values of d_i, we can identify the states of convergence or divergence of relative per capita environmental indicators.

Scenario 1: $-0.5 < d \leq 0$. This is a short memory process, i.e., there is "fast convergence" or "short memory convergence."
Scenario 2: $0 < d < 0.5$. This is a long memory but still stationary process, and also a convergent process with slowly (or smoothly) decaying speed. In this case, a country's environmental indicators might undergo a long period of convergence toward a common long-term trend.
Scenario 3: $0.5 < d < 1$. This is a long memory process, which is non-stationary but still has the property of mean reversion. In this case, the process is characterized by a high degree of persistence. Therefore, the impact of the distant past will still have lasting effects on the present.
Scenario 4: $d \geq 1$. This is a unit root or divergent process. In this case, any initial shock has a significant impact on the future, and the process is unlikely to return to the mean at some point in the future.

The testing approach to the fractionally integrated order allows a richer classification of convergence states, so that stationary convergence and non-stationary convergence but with mean reversion can be distinguished. Another characteristic of this classification scheme is that it can take into account the lasting impact of the previously invoked shock on the present, or the rapid attenuation without any impact on the present, and more importantly, it can take into account the states in between two extremes mentioned above. The simple $I(0)|I(1)$ dichotomy could not capture these states because the standard unit root test can only considers two extreme cases, i.e., persistence or no persistence at all.

Then, we use local Whittle likelihood estimation method (LW) and various improved versions to estimate the parameter of fractional integration d of Equation (2). The advantage of this method is that the estimator is not affected by the non-stationarity of the short-term process. In order to judge whether a per capita environmental indicator is an unit root process or a mean-reverting and long memory process with finite or inflated variance (depending on the estimated $\hat{d}$ value), we conduct hypothesis test of the null hypothesis $H_0 : \hat{d} = d_0$ and the alternative hypothesis $H_1 : \hat{d} \neq d_0$. If the hypothesis holds that per capita relative environmental indicators are mean-reverting, there is enough evidence to prove that these indicators are stochastically convergent processes.

After applying LW, ELW, two-stage detrended ELW (2ELWd) and the 2ELWd method with breakpoints removed to obtain an estimate of the order of the fractional integration $\hat{d}$ of the individual environmental indicator series, we use the Robinson (33)'s LM test framework to test the hypothesis of the environmental indicator series of each country as shown in Equation (4):

$$H_0 : d = d_0 \quad (4)$$

In this hypothesis test, the null hypothesis is the existence of unit root. The alternative is that the per capita environmental indicators have long memory characteristics. Robinson (33) proposed the following statistics and proved that the standard normal distribution is satisfied under certain regularity conditions:

$$\hat{R} = \left(\frac{n}{\hat{A}}\right)^{1/2} \frac{\hat{a}}{\hat{\sigma}^2} \sim^d N(0,1), n \to \infty \quad (5)$$

where,

$$\hat{A} = \frac{2}{n}\left[\sum_{j=1}^{n-1}\psi\left(\lambda_j\right)^2 - \sum_{j=1}^{n-1}\psi\left(\lambda_j\right)\hat{\varepsilon}\left(\lambda_j\right)' \times \left(\sum_{j=1}^{n-1}\hat{\varepsilon}\left(\lambda_j\right)\hat{\varepsilon}\left(\lambda_j\right)'\right)^{-1} \times \sum_{j=1}^{n-1}\hat{\varepsilon}\left(\lambda_j\right)\hat{\varepsilon}\left(\lambda_j\right)'\right] \quad (6)$$

$$\hat{a} = \frac{-2\pi}{n}\sum_{j=1}^{n-1}\psi\left(\lambda_j\right)g\left(\lambda_j,\hat{\tau}\right)^{-1}I\left(\lambda_j\right) \quad (7)$$

$$\psi\left(\lambda_j\right) = \log\left(2\sin\frac{\lambda_j}{2}\right);\ \hat{\varepsilon}\left(\lambda_j\right) = \frac{\partial}{\partial\tau}\left[\log g\left(\lambda_j;\hat{\tau}\right)\right];\ \lambda_j = \frac{2\pi j}{n}.$$

$I\left(\lambda_j\right)$ represents the periodogram of $\hat{u}_t$ as follows:

$$I\left(\lambda_j\right) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n}\hat{u}_t e^{i\lambda_j t}\right|^2 \quad (8)$$

where $\hat{u}_t$ is the least squares residual obtained from Equation (2), g is a known function from the spectral density:

$$f\left(\lambda_j;\tau\right) = \frac{\sigma^2}{2\pi}g\left(\lambda_j;\tau\right) \quad (9)$$

Robinson (33)'s test framework relies on some specific assumptions imposed on the short memory component u_t. If it is a white noise, $g \equiv 1$; If it is a process that follows as the form of AR (1) $u_t = \tau u_{t-1} + \eta_t$, $g\left(\lambda_j;\tau\right) = \left|1-\tau e^{i\lambda_j}\right|^{-2}$, and σ^2=var (η_t). The parameter considered is the long memory parameter d, which shows whether a time series process is long memory. Because that $d < 1$ means the relative per capita environmental indicators converge, the null hypothesis is set as follows in this study:

$$H_0 d = d_0 = 1. \quad (10)$$

This study uses the data set of the BRICS countries from 1961 to 2017 for empirical analysis. Obviously, the data set is a small sample ($n = 57$), so it is necessary to calculate the empirical distribution of various local Whittle estimators. When the sample size $n = 57$ and the truncation parameter α is in the interval [0.65, 0.8], we will give the accurate distribution of the memory parameter d estimators of LW, ELW, 2LWD and the detrended 2LWD method, and then construct the confidence interval of the estimators.

Convergence Pseudo-Long Memory Test of Relative Environmental Indicators per Capita

That relative environmental indicators are long memory processes indicates that it takes a long time for these indicators to approach convergent paths. However, although a time series (as mentioned above y_{it}) is a long-term memory process, it is not clear whether it is indeed a long memory process or a short memory and stationary one with a mean shift (showing the characteristics of long-term memory processes). Although these two cases show that the sequences of relative environmental indicators per capita converge, it is slower for the stationary processes with mean shifts to approach convergence than the real long memory processes. If their means shift toward the upward direction of the sequences, these indicators converge more slowly; If means shift downwards, these sequences converges faster. After obtaining evidence that these indicators are long memory I (d) or infinite memory I (1) processes, this paper will continue to determine whether the potentially slow convergence or lack of convergence is real or the result of a structural change in the mean of an otherwise stationary process.

This paper introduces a test method proposed by Mayoral (2) to solve this spurious long memory problem. If per capita environmental indicators are pseudo-long memory processes, it means that these processes are stationary, but at a certain time point, a structural fracture occurs in the opposite direction of the sequence trend, indicating that the per capita environmental indicator would reach convergence after a shorter time. These two types of models fitting the two processes can be nested together and expressed as Equation (11) below. This paper assumes that the relative environmental indicator data set $y_1, y_2, \cdots, y_n$ are generated from the model listed as follows:

$$y_t = \beta' \cdot Z_t + \delta \cdot V_t(\omega) + x_t \quad (11)$$

Where Z_t is a deterministic component, and $V_t(\omega)$ can be a constant or a trend function. It is used to judge whether the time series has undergone structural changes. The observations of the time series are Z_{t-k} after the break point appears, while they are evaluated as zeros before the break point appears. Meanwhile, the random term x_t is defined as:

$$(1-L)^d x_t = u_t,\ \ \varepsilon_t \sim N\left(0,\sigma^2\right) \quad (12)$$

Based on this data-generating structure, the hypothesis testing should set the following null and alternative hypothesis:

$$\begin{cases} H_0 : d \in D_0,\ \ \delta = 0, D_0 \subset (0.5, 1.5) \\ H_1 : d = 0,\ \ \delta\ \ noconstraint \end{cases} \quad (13)$$

To test the above hypotheses, Mayoral (2) proposed a semi-parametric test method, and constructed a Most Powerful Invariant (MPI) statistics by comparing the log-likelihood ratio under the null hypothesis H_0 and the alternative one H_1. This consistent statistics $R_b^f\left(\hat{d}_n\right)$ is of the form as follows:

$$R_b^f\left(\hat{d}_n\right) = n^{1-2\hat{d}_n}\left(\frac{\hat{\lambda}^2}{\hat{\gamma}_0}\right)^{-1}\frac{\inf_{\omega\in\Omega}\left(y - Z\cdot\tilde{\beta} - V(\omega)\tilde{\delta}\right)'\left(y - Z\cdot\tilde{\beta} - V(\omega)\tilde{\delta}\right)}{\left(\Delta^{\hat{d}_n}y - \Delta^{\hat{d}_n}Z\cdot\hat{\beta}\right)'\left(\Delta^{\hat{d}_n}y - \Delta^{\hat{d}_n}Z\cdot\hat{\beta}\right)} \quad (14)$$

where the order $\hat{d}_n$ of fractional integration is the uniform estimator obtained by ELW and other methods. As a semi-parametric test method, the statistics should be evaluated by resorting to the ordinary least square method (OLS) to estimate the parameters in Equation (11) under H_0 or H_1 hypothesis, respectively. All the parameters of the numerator, including $\tilde{\beta}$ and $\hat{\delta}$ on the right side of the Equation (14) are estimated when the null hypothesis H_0 holds, while $\tilde{\beta}$ and $\hat{\delta}$ in the denominator are estimated under the alternative hypothesis H_1. At the same time, under the null hypothesis H_0, $\hat{\gamma}_0$ represents ε_{it} variance estimator. $\omega = k/n$ is the ratio of the number of sub-sample observations before the possible break point to the total sample. Considering that allowing the break point to appear in the whole interval (0, 1) will cause the test to have a very low power potentially, we adopt the restricted interval $\Omega = [0.15, 0.85]$ suggested by Andrews (34). In addition, the parameter λ^2 in Equation (14) can be obtained from the following formula:

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{i=1}^{q} k\left(\frac{i}{q}\right)\hat{\gamma}_i \tag{15}$$

where γ_i is the i autocovariance function of ε_t, and $k(\cdot)$ is the Bartlett kernel function. q is the bandwidth parameter. According to the method of Newey and West (35), we take $q = 4 \times (n/100)^{2/9}$ as the initial value and select the bandwidth automatically decided by data. Mayoral (2) proved that $R_b^f\left(\hat{d}_n\right)$ is a uniform statistics.

For the initial estimate of fractionally integrated order of a single sequence and the test under the null hypothesis that $\hat{d} = 1$, the Monte Carlo and Bootstrap critical values of the statistics $R_b^f\left(\hat{d}_T\right)$ are given by means of 20,000 repetitions of the simulation of the time series data with sample size of 57. If the value of the statistic $R_b^f\left(\hat{d}_n\right)$ is less than a critical value at a confidence level, this is evidence to reject the null hypothesis, which means that the time series is pseudo long-memory.

Identification of Time Point of Structural Mutation of Convergence Path of Relative Environmental Indicators per Capita

Although the test method proposed by Mayoral (2) has good test effect in identifying true long- memory (including infinite-memory, i.e., unit root processes) processes from pseudo long-memory ones, it is not a accurate break point detection method (23). This means that it can only identify whether a process is a long-term memory (non-stationary process) or a stationary process with a mean shift, but cannot identify the exact location of the break point in the case of "pseudo long-memory processes." Therefore, it is necessary to identify the specific positions of all break points in the sequence by perfecting the above-mentioned test method. For this purpose, this study uses the null hypothesis H_0 proposed by Berkes et al. (1), shown by the Equation (16):

$$\begin{cases} y_t = \mu + x_t, \ 1 < t < k^* \\ y_t = \mu + \Delta + x_t \ k^* + 1 < t < n \end{cases} \tag{16}$$

where k^* is a point in time when the mean-shift happens, and μ and $\mu + \Delta$ is the unknown mean and shift of the potential data generated process respectively. The sequence $\{x_t\}$ is a fourth-order stationary process with a mean of zero. This study uses the non-parametric CUSUM statistics proposed by Berkes et al. (1) and Aue et al. (36) to test H_0 listed below.

H_0: The observations $y_1, y_2, \cdots, y_n$ follow the mean-shift process (21) and x_t is the stationary process with a weak auto-correlated structure.

H_1: The observations $y_1, y_2, \cdots, y_n$ are a long memory process, i.e. their potential model setting satisfies Equation (12).

For defining hypothesis test statistics, we apply the main ideas of Berkes et al. (1) to introduce a break point that occurs at time point $\hat{k}^*$, and estimate it by the following equation:

$$\hat{k}^* = \min\left\{k : \max_{1<i<n}\left|\sum_{1\le j\le n} y_j - \frac{i}{n}\sum_{1\le j\le n} y_j\right| = \left|\sum_{1\le j\le k} y_j - \frac{k}{n}\sum_{1\le j\le n} y_j\right|\right\} \tag{17}$$

Since the identification program is designed for a single break point, we take the minimum $\{\cdot\}$ in the algorithm for detecting possible breakpoints to obtain the position of the strongest mean shift point. Therefore, the entire sample is divided into two sub-samples dependent on the break point $\hat{k}^*$. In each sub-sample, the T_n statistic can be constructed according to the following two formulas:

$$T_{n,1} = \frac{1}{s_{n,1}}\hat{k}^{*-\frac{1}{2}} \max_{1\le k\le \hat{k}^*}\left|\sum_{1\le i\le k} y_i - \frac{k}{\hat{k}*}\sum_{1\le i\le n} y_i\right| \tag{18}$$

$$T_{n,2} = \frac{1}{s_{n,2}}\left(n - \hat{k}^*\right)^{-\frac{1}{2}} \max_{\hat{k}^*<k\le n}\left|\sum_{\hat{k}^*<i\le k} y_i - \frac{k-\hat{k}^*}{n-\hat{k}^*}\sum_{\hat{k}^*<i\le n} y_i\right| \tag{19}$$

In the above equations, the CUSUM statistics is standardized by long-term standard deviations $s_{n,1}$ and $s_{n,2}$ respectively. $s_{n,1}$ is the standard deviation calculated based on the first piece of sample data $y_1, y_2, \cdots, y_{\hat{k}}$; $s_{n,2}$ is the standard deviation calculated based on the second part of sample data $y_{\hat{k}+1}, y_{\hat{k}+2}, \cdots, y_n$. There are several different methods to estimate the long-term variance s_n^2. For example, if the series $\{y_t\}$ has a serial correlation structure, it can be estimated according to s_n^2 as follows:

$$s_n^2 = \hat{\gamma}_0 + 2\sum_{1\le j\le q(n)} \omega_j\left(q(n)\right)\hat{\gamma}_j \tag{20}$$

The $\hat{\gamma}_i$ can be calculated as follows:

$$\frac{1}{n}\sum_{1\le i\le n-j}\left(y_t - \bar{y}_n\right)\left(y_{t+j} - \bar{y}_n\right) \tag{21}$$

Where $\omega_j(\cdot)$ is a weighting kernel function, and $q(n)$ is the bandwidth of the kernel function $\omega_j(\cdot)$. Andrews (34) discussed

seven different types of kernel functions that can be used to calculate the HAC estimator. We select the "Bartlett" kernel function $\omega_j(q) = 1 - j/(q+1)$ according to Berkes et al. (1) and Aue et al. (36) with the optimal bandwidth used by Newey West in the selected kernel function. Once the T_n statistics of the two sub-samples are obtained, the following M_n statistics can be constructed:

$$M_n = \max\left\{T_{n,1}, T_{n,2}\right\} \tag{22}$$

T_n can be used to discriminate whether the sample has been generated by a short or long-range dependent stationary process. Hence, if we split the sample at time $\hat{k}^*$, which is close to the true change-point k^* asymptotically we can assume that $y_1, y_2, \cdots, y_{k^*}$ and $y_{k^*+1}, y_{k^*+2}, \cdots, y_n$ are samples from a stationary sequence with a constant mean. Thus, we can use $T_{n,1}$ and $T_{n,2}$ to test if the samples $y_1, y_2, \cdots, y_{k^*}$ and $y_{k^*+1}, y_{k^*+2}, \cdots, y_n$ have been generated by a short-range or long-range dependent process. The CUSUM statistic compares the sample mean of the first k^* observations with the sample mean of the last $(n-k)$ observations. If we assume a change in the mean at time k^*, the absolute value of the difference of the means should be large for $k = k^*$.

Berkes et al. (1) derived the following asymptotic distributions of the statistics M_n under the null hypothesis:

$$M_n \to^d \max\left\{\sup_{0\le t\le 1}|B^{(1)}(t)|, \sup_{0\le t\le 1}|B^{(2)}(t)|\right\} \tag{23}$$

where $B^{(1)}(t)$ and $B^{(2)}(t)$ is a Brown bridge corresponding to two sub-samples, respectively. If the relative environmental indicator per capita is weakly dependent (there is no change in mean), the statistics M_n will converge to the upper bound of the Brown bridge. However, if an environmental indicator fellows a mean shift or a long memory process, $M_n \to \infty$. The latter case is referred to as a false rejection of the null hypothesis that there is no change in mean. Because the dataset used here is not large enough to achieve an approximately asymptotic distribution. Therefore, we calculate the asymptotically critical values under H_0 at the significance level of 10, 5 and 1% respectively. If M_n exceeds the critical value at a significance level, the null hypothesis cannot be rejected, and the breakpoint identified at the corresponding significance level is reasonable.

This paper follows the procedure shown in **Figure 1** to test whether ecological indicators are convergent in BRICS countries.

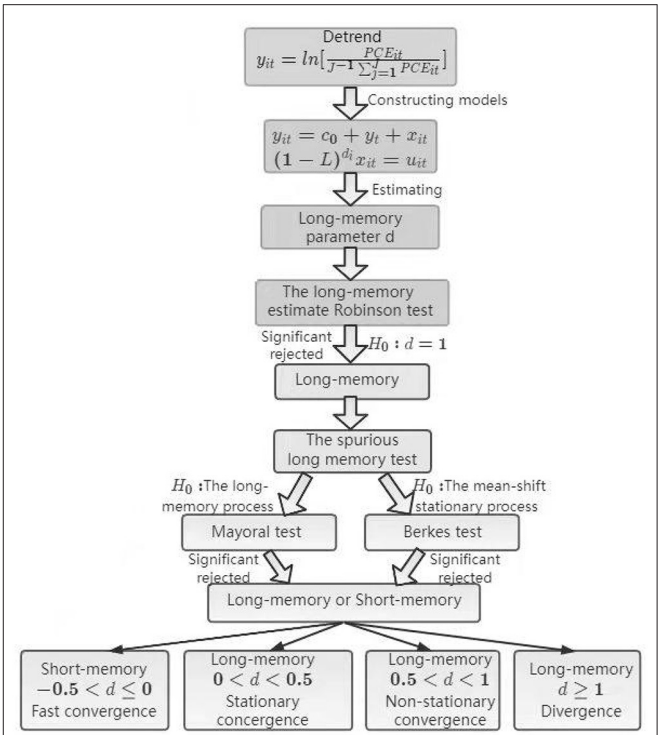


FIGURE 1 | A flow diagram for testing the long-memory of ecological indicators.

DATA

The data used in this paper are from National Footprint Accounts (NFAs). NFAs provides global and national annual accounting of biocapacity (BC) and ecological footprint (EF). This paper collects annual data of EF and BC per capita at the national level. The data range is from 1961 to 2017, with a total time span of 57 years. **Figure 2** shows the trend of per capita ecological footprints and ecological carrying capacities of the BRICS countries (China, Brazil, India, Russia and South Africa) and the world. During the sample period, EFs are larger than the quantities enough to ensure sustainable development in BRICS and the world. In other words, the existing productive land and other resources are insufficient to meet the needs of human food, housing and CO_2 absorption. During this period, around the worldwide, per capita EF increases by nearly 120%. BC per capita decreases by 50% during the same period. In 2017 the ecological deficit fell to 5.26 gha (a measure unit of ecological footprint and ecological carrying capacity), while EF almost doubled. This shows that human activities have been imposing great pressures on the nature. Although the ecological footprint can maintain the trend of not decreasing, the ecological carrying capacity decreases rapidly and the ecological deficit continues to increase every year. All three environmental indicators reveal that globally, human beings consume more resources with ecological productivity than available. For example, from 1961 to 2017, the ecological carrying capacity per capita decreases by 41.4%, from 3.12 gha to 1.83 gha per capita. This fact means that resources available per capita have been decreasing and more people are competing for limited resources (37).

Figure 1 shows that the world's per capita EF was stable in the 1980s and 1990s, and has shown an upward trend since the beginning of the twenty-first century, while per capita BC decreased gradually. Since 1961, the world EF per capita has been steadily increasing at a rate of 2.1% per year, almost tripling from 4.62 gha to 5.24 gha in 2017. **Figure 2** also shows the trend of the per capita EFs and ecological carrying capacities of BRICS countries over time. In most countries, the

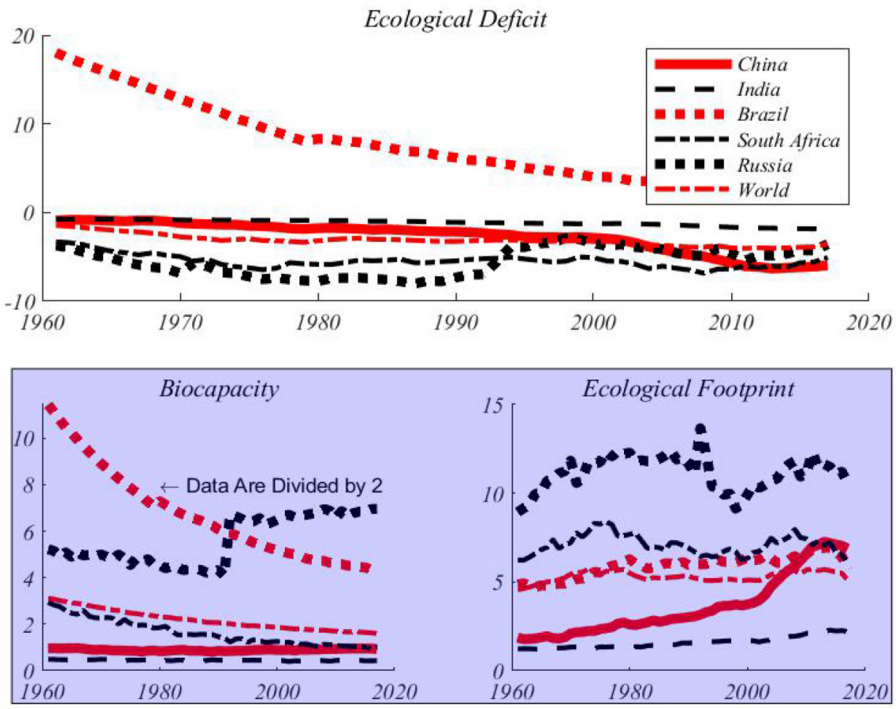


FIGURE 2 | Per capita environmental indicators for BRICS and the world (gha). Russian data before 1989 are replaced with data from the former Soviet Union. So you can see from the fifth chart that there is a jump in 1989.

per capita EFs continue to increase, while the per capita BCs are gradually decreasing or remain unchanged. With the rapid economic development of the BRICS countries, the problem of environmental pollution is becoming more and more serious. In 2017, all BRICS countries except Brazil were in ecological deficit. Compared with previous years, this deficit were very high in most of these countries. Therefore, it is necessary to study the stochastic evolution path of ecological indicators in BRICS. The EFs increase significantly in BRICS except for South Africa, especially after 1985. The per capita EFs have risen sharply in China and India, the biggest economies, indicating that these two countries are rapidly consuming natural resources. Although the trend of increase in per capita EF in Russia is not obvious, the level of per capita EF in Russia is higher than the sum of EFs in China and India. Although Brazil is the only BRICS country with an ecological surplus, its per capita BC is also declining at the fastest rate among BERICS, reaching the brink of an ecological deficit in 2017. Obviously, the level of economic development is closely related to the per capita EF. On the other hand, South Africa's ecological carrying capacity is obviously declining rapidly. In the past decade, China and South Africa have undergone higher ecological deficits than other countries. This means that the two countries with higher per capita incomes have put more pressures on the environment and faced a bigger problem of environmental deterioration. Environmental sustainability requires that the consumption of resources by the present generation cannot pose a threat to the needs of future generations. The imbalance between ecological footprints and ecological carrying capacities means that the present society is using the resources of future generations to live. In order to prevent the trend of further deterioration of the ecological deficit, the close cooperation between countries is required to take actions to improve the modes of production of enterprises and consumption patterns of human beings. **Table 1** shows the descriptive statistics of BRICS and the world environmental indicators. From these summary statistics, we can see that there are also great differences in environmental indicators among BRICS. Russia has the highest per capita EF, while Brazil has the highest per capita BC.

From **Figure 2**, it is also clear that the CFs and BCs are persistent, BCs downwards and CFs upwards. With respect to most of series, the processes might be not mean-reverting, and exhibit long memory properties presumably from the argument proposed by Granger and Joyeux (38) that the aggregation of first order Markov processes leads to a long memory process.

EMPIRICAL ANALYSIS

Analysis of Convergence and Divergence of per Capita Ecological Indicators (EF and BC)

Convergence and Divergence of per Capita Ecological Indicators

Tables 2, **3** report the memory parameter estimates of the fractionally integrated processes estimated using LW, ELW

TABLE 1 | Summary statistics of per capita environmental indicators in BRICS and the world.

Countries		China	India	Brazil	South Africa	Russia	World
Mean	EF	3.635	1.588	5.889	7.097	11.153	5.306
	BC	0.896	0.448	13.358	1.638	5.596	2.182
Median	EF	2.974	1.557	6.033	7.065	11.241	5.268
	BC	0.892	0.449	12.50	1.544	5.100	2.074
Maximum	EF	7.273	2.289	6.931	8.394	13.671	5.734
	BC	0.973	0.492	22.892	2.935	6.963	3.116
Minimum	EF	1.779	1.199	4.707	6.095	8.037	4.564
	BC	0.819	0.394	8.609	0.958	4.083	1.598
Std. Dev.	EF	1.740	0.322	0.603	0.564	0.987	0.286
	BC	0.044	0.021	3.993	0.553	1.033	0.440
Skewness	EF	0.959	0.792	−0.515	0.417	−0.319	−0.440
	BC	0.183	0.005	0.773	0.702	0.068	0.568
Kurtosis	EF	2.598	2.517	2.423	2.601	2.780	2.730
	BC	1.914	2.913	2.547	2.311	1.251	2.141
JB	EF	9.128	6.516	3.305	2.029	1.082	2.009
	BC	3.118	0.018	6.158	5.805	7.312	4.812
P Value	EF	0.019	0.033	0.099	0.228	0.492	0.232
	BC	0.109	0.500	0.037	0.040	0.027	0.054

(1) The last three lines represent the P values corresponding to the JB statistics. (2) "—" represent missing data.

and the ELW methods with detrended (or with removed breakpoints) for each environmental indicator sequence. The integration order of each environmental indicator in all countries is greater than 0, indicating that all sequences exhibit long memory characteristics. For each estimated memory parameter, a corresponding 95% confidence interval is provided. According to the confidence interval range of fractionally integrated proesses' memory parameters, it can be found that, in general, the confidence intervals of the memory parameter estimates of per capita ecological footprints in all countries have low bounds higher than 0.5, indicating that these indicators may be non-stationary. However, the low bounds of the confidence intervals of the per capita BC memory parameter estimates of BRICS are less than 0.5, which means that the per capita BCs may be stationary. For most countries, the memory parameter estimates of the ecological footprints per capita are greater than 1, which means that they will not converge to their common average of BRICS. Therefore, it is possible to reject both pure stationarity and unit root. In some cases, the effect of a given random shock will be temporary, but the sequence tends to return to its trend and converges slowly at a slower rate than a pure stationary sequence. The memory parameter of per capita BC sequences in BRICS are <1, indicating that these sequences are mean reverting (i.e., convergence). These insights can be analyzed in detail as follows.

As for testing per capita ecological indicators' convergence, the Monte Carlo critical values and bootstrap critical values of the Robinson test make the consistent conclusion, indicating that the per capita ecological footprint sequences of China and India are divergent; The per capita ecological footprints of Russia and South Africa are convergent, but the processes of convergence are quite slow; The per capita ecological footprint of Brazil shows a very weak convergence dynamics. All countries expect for South Africa have memory parameters of per capita carbon footprint series significantly >1, which means that the per capita carbon footprints of these four countries are far from convergence. China's relative ecological footprint per capita has the highest persistence ($\hat{d} = 1.190$); Russia's per capita relative ecological footprint has the lowest persistence ($\hat{d} = 0.643$), indicating that Russia's per capita carbon dioxide emissions are serious. The confidence interval for the estimated fractionally integrated process parameter (long-memory parameter) is very narrow and generally located within a positive interval range. For the relative ecological footprints per capita, the upper bound of the confidence intervals are generally >0.5, indicating that the ecological footprints are non-stationary and some are even mean-reverting, but exhibit long memory characteristics. The two environmental indicators in other countries do not have the feature of mean reversion and are not convergent. For South Africa and Russia, the inference results, i.e., convergence or divergence, are very consistent, no matter Monte Carlo or bootstrap approach is used to statistically infer. Specifically, the range of estimates $\hat{d}$ is between 0.49 and 0.72 (statistically always different from 1), indicating long memory and mean-reverting (convergent) processes, but not covariance stationary processes. Therefore, the per capita relative ecological footprints of China, India and Brazil are not convergent, while those indicators of South Africa and Russia are convergent, although the convergence will take quite long time ($\hat{d} = 1.190$).

However, as for the ecological carrying capacities, it is another story. The upper bounds of the memory parameters $\hat{d}$ of the BC sequences per capita in most of BRICS countries are greater than 0.5, but less than 1, which indicates that these sequences

TABLE 2 | Long memory parameter estimates of per capita EFs in BRICS.

	LW	ELW	2ELWd	De-breakpointed 2ELWd
China	0.980	1.205	1.190	1.116
	[0.973 1.172]	[0.761 1.412]	[0.799 1.359]	[0.696 1.302]
$H_0 : \hat{d} = 1$	−26.567	−48.058	−47.157	−41.832
MC C V	17.959(+)	13.787(+)	24.458(+)	16.041(+)
Bootstrap CV	4.836(+)	4.809(+)	4.877(+)	4.185(+)
India	0.889	1.184	1.182	1.002
	[0.726 1.157]	[0.792 1.351]	[0.772 1.357]	[0.582 1.167]
$H_0 : \hat{d} = 1$	58.948	4.140	−23.313	15.711
MC CV	8.495(+)	11.109(+)	12.052(+)	15.041(+)
Bootstrap CV	3.656(+)	3.205(+)	0.364(+)	3.661(–)
Brazil	0.762	1.148	1.143	0.932
	[0.579 1.068]	[0.766 1.300]	[0.743 1.336]	[0.480 1.114]
$H_0 : \hat{d} = 1$	−0.856	−1.777	−0.858	−1.238
MC CV	11.436(+)	12.304(+)	9.135(+)	13.402(+)
Bootstrap CV	3.957(+)	3.973(+)	3.966(+)	3.926(+)
South Africa	0.880	0.934	0.892	0.853
	[0.707 1.162]	[0.524 1.14]	[0.459 1.062]	[0.853 1.009]
$H_0 : \hat{d} = 1$	3.277	−1.893	2.943	4.072
MC CV	13.574(+)	12.041(–)	11.442(–)	11.391(–)
Bootstrap CV	3.078(+)	3.049(–)	3.076(–)	3.067(–)
Russia	0.823	0.924	0.869	0.643
	[0.632 1.125]	[0.561 1.067]	[0.477 1.017]	[0.202 0.737]
$H_0 : \hat{d} = 1$	53.153	28.168	40.206	155.513
MC CV	8.115(–)	13.451(–)	11.634(–)	13.237(–)
Bootstrap CV	3.335(–)	3.351(–)	3.343(–)	3.253(–)

In this paper, Bootstrap percentile -t confidence intervals are used for evaluating the estimates of memory parameters, at 95% confidence level. (+) represents rejecting null hypothesis; while (–) represents accepting null hypothesis. We make inference dependent on 5% significant level. With respect to the Robinson statistics represented by Equation (4), empirical critical values are generated by 2000 Monte Carlo repetitions and 2000 Bootstrap resamples, respectively, the corresponding contains of following tables are the same as this table.

TABLE 3 | Long memory parameter estimates of per capita BCs in BRICS.

	LW	ELW	2ELWd	De-breakpointed 2ELWd
China	0.889	0.752	0.844	0.663
	[0.775 1.008]	[0.369 0.752]	[0.395 0.998]	[0.192 0.787]
$H_0 : \hat{d} = 1$	231.241	608.571	326.933	1068.545
MC CV	18.627 (–)	14.856 (–)	14.382 (–)	20.292 (–)
Bootstrap CV	3.757 (–)	3.835 (–)	3.801 (–)	3.776 (–)
India	0.824	0.811	0.874	0.668
	[0.618 0.987]	[0.352 0.895]	[0.461 1.041]	[0.257 0.773]
$H_0 : \hat{d} = 1$	366.628	396.608	266.935	940.048
MC CV	22.274 (–)	9.843 (–)	11.593 (–)	11.695 (–)
Bootstrap CV	4.080 (–)	4.905 (–)	4.051 (–)	4.056 (–)
Brazil	0.903	1.047	1.017	0.290
	[0.835 1.067]	[0.646 1.173]	[0.618 1.110]	[-0.236 0.383]
$H_0 : \hat{d} = 1$	152.153	−112.29	−68.358	155.375
MC CV	9.396 (–)	17.748(+)	13.938(+)	11.759 (–)
Bootstrap CV	3.907 (–)	3.956(+)	3.965(+)	3.946 (–)
South Africa	0.824	0.733	0.586	0.018
	[0.605 1.046]	[0.351 0.745]	[0.172 0.696]	[-0.375 0.196]
$H_0 : \hat{d} = 1$	48.674	97.146	290.986	169.625
MC CV	14.134 (–)	13.445 (–)	19.715 (–)	8.285 (–)
Bootstrap CV	2.919 (–)	2.959 (–)	2.933 (–)	2.958 (–)
Russia	0.907	0.948	0.928	0.287
	[0.819 1.113]	[0.534 1.064]	[0.380 1.024]	[-0.273 0.421]
$H_0 : \hat{d} = 1$	3.642	1.860	2.667	319.372
MC CV	10.101(+)	9.221(+)	10.610(+)	9.919(+)
Bootstrap CV	3.512 (–)	3.494(+)	3.459(+)	3.514(+)

are mean-reverting, but of not stationary. These results show that in all BRICS countries the BC sequences are convergent, with quite different convergent rates. In China and India, the BC sequences have the longest memory degree, and the estimated memory parameters are $\hat{d} = 0.668$ and $\hat{d} = 0.663$, respectively, which means the two countries have the slowest convergent paths. However, the memory parameter of South African per capita BC sequence is that $\hat{d} = 0.018$, in other words no long memory characteristic, i.e. converging quickly. The memory parameter estimate of the BC sequence of Russia is less than 0.5 ($\hat{d} = 0.287$). The analysis above means that the per capita BC is a stationary process, but a long memory process, converging in a relatively fast rate. From the policy perspective, as long as the government adopts environmental protection and resource use policies designed appropriately, the decline trends of the per capita ecological carrying capacities of BRICS can be reversed and eventually converge. It is not as stubborn as the worsening trend of the ecological footprints, which have put tremendous pressures on the governments of BRICS.

Statistical Test of Convergence and Divergence of per Capita Ecological Indicators (EF and BC)

We now turn our attention to investigating the finite sample simultaneous coverage probability and effectiveness of the critical intervals for the Robinson (33)'s tests. This kind of graphic is very useful for choosing among critical values that have reasonable coverage distortions and effectiveness improvements: it permits to make arbitrage between the coverage distortion and the true effectiveness for each critical values and then to chose the most appropriate. More specifically, we do both the IID and Monte Carlo and bootstrap sampling respectively for 2,000 independent replications with $n = 64$. In each replication, we calculate uncorrected marginal intervals. We then plot the point estimate of the coverage and effectiveness along with simultaneous confidence regions in **Figures 3**, **4**.

We consider coverage plots as suggested by de Peretti and Siani (5) defined as follows. For a nominal coverage x the coverage is calculated as $x_s = 1 - 2\min\{pv, 1 - pv\}$, where pv is equal to

$$pv = \frac{1}{B}\sum_{b=1}^{B} I\left(\hat{\tau} \leq \tau_0\right). \tag{24}$$

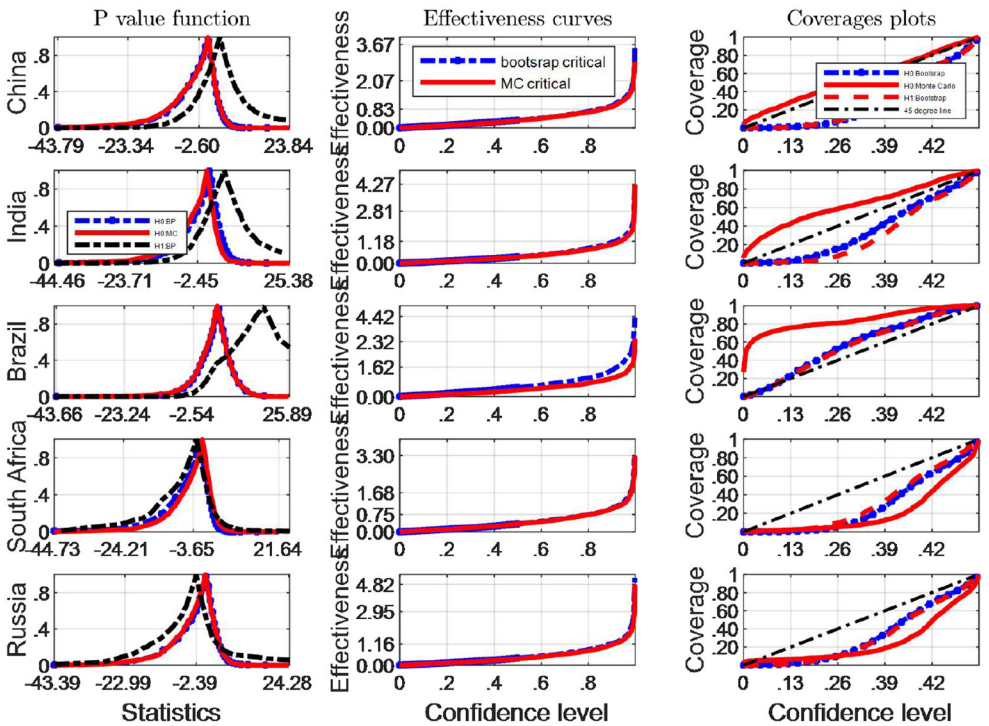


FIGURE 3 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Robinson test using EF indicators.

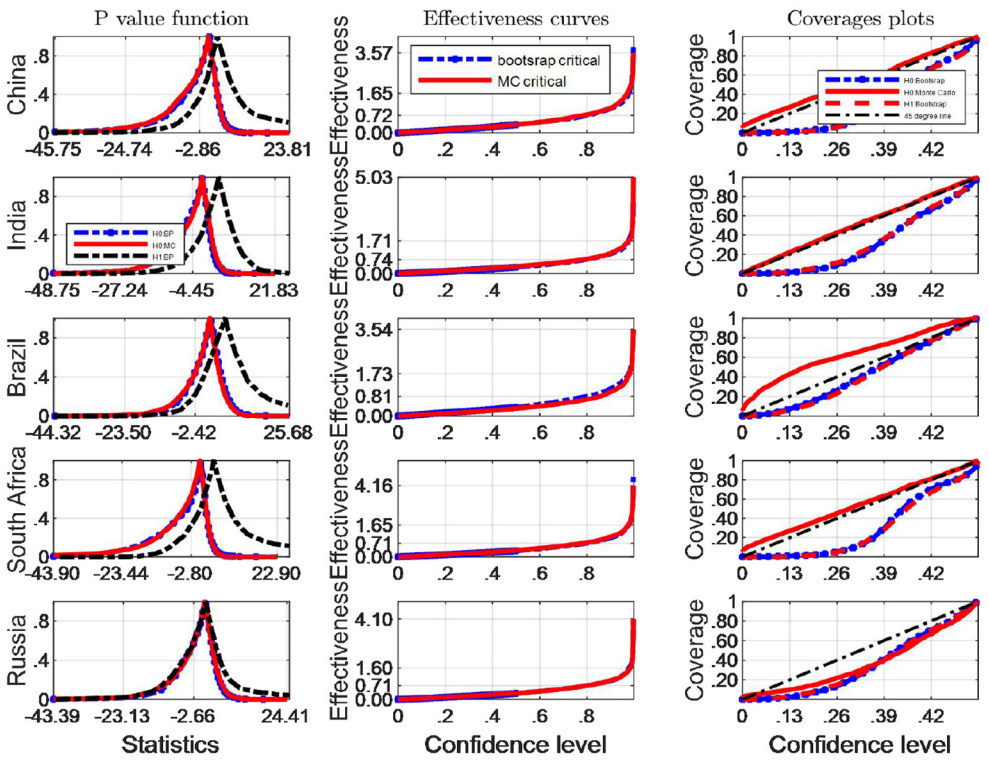


FIGURE 4 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Robinson test using BC indicators.

TABLE 4 | Pseudo-long memory Mayoral test of relative per capita environmental indicators in BRICS.

	LW		ELW		De-trented (breakpoint) 2ELWd	
	EF	BC	EF	BC	EF	BC
China						
Mayoral statistics	0.069	0.081	0.0112	0.255	0.023	0.530
MC CV	0.051	0.072	0.0095	0.214	0.018	0.438
Bootstrap CV	0.039	0.073	0.009	0.201	0.016	0.393
India						
Mayoral statistics	0.165	0.142	0.150	0.158	0.068	0.510
MC CV	0.085	0.105	0.007	0.116	0.033	0.386
Bootstrap CV	0.073	0.116	0.010	0.128	0.034	0.368
Brazil						
Mayoral statistics	0.908	0.170	0.042	0.051	0.244	13.891
MC CV	0.260	0.066	0.012	0.020	0.065	9.158 (–)
Bootstrap CV	0.189	0.067	0.013	0.025	0.055	7.924 (–)
South Africa						
Mayoral statistics	0.106	0.095	0.067	0.210	0.133	64.882
MC CV	0.068	0.085	0.043	0.184	0.085	74.315(–)
Bootstrap CV	0.080	0.119 (–)	0.054	0.234 (–)	0.096	76.952 (–)
Russia						
Mayoral statistics	0.154	0.135	0.067	0.095	0.562	8.83
MC CV	0.115 (–)	0.057 (–)	0.051	0.040	0.524	9.500 (–)
Bootstrap CV	0.119 (–)	0.065 (–)	0.058	0.049	0.468	8.238

With respect to Mayoral statistics shown by Equation (14), empirical critical values are obtained by 2,000 repetitions of Monte Carlo simulation and 2,000 repetitions of bootstrap resample respectively.the corresponding contains of following tables are the same as this table.

where pv is the estimated probability distribution function at the true Robinson (27)'s statistics $\hat{\tau}$, using any of the bootstrap and Monte Carlo strategies described above. τ_0 is the correspoding critical value, and B simulation repetitions. The closer the coverage line is to 45 degree line the better the strategy used to approximate the distribution of Robinson statistics performs. The third column of **Figures 2**, **4** presents the coverage plots of the interval based on inverting (single) bootstrap tests for the five BRICS countries. Not only is the frequency of coverages important to assess the performance of a CI but its length is also a relevant measure of its effectiveness [see (5)]. If two CIs have the same confidence level the shorter is preferable, indicating higher precision in the estimation. The second column of **Figures 3**, **4** presents the confidence level-effectiveness curves of the interval based on inverting (single) bootstrap tests based on the parameter estimates of the five counties. On the basis of confidence level-effectiveness curves, two critical values seem to have the same effectiveness (expect for Brazil) and to be dominated differently. Following the results on **Figures 3**, **4**, the Bootstrap method has the most satisfactory "true" effectiveness for Brazil.

As is analyzed above, **Figure 3** shows clearly that MC critical values should be resorted to judge the Robinson tests for EF indicators, and the confidence interval based on inverting bootstrap tests is by far the best method on the basis of coverage accuracy criterion in China and India. However, the bootstrap critical values should be used to infer Robinson tests in other BRICS countries since these coverage curves approaches 45 degree lines closer, and their confidence intervals are narrower.

It is obvious that the Bootstrap critical value does not perform well for judging the statistic hypothesis test of BC indicators in almost all BRICS countries (shown by **Figure 4**) since the MC plot is closer to the 45 degree line. Consequently, in these countries MC critical values should be employed to infer Robinson test. However, in Brazil, Bootstrap critical values are more appropriate tools to decide the inference of Robinson test at a significance level.

Pseudo-Long Memory Test (Mayoral Test) for Per Capita Environmental Indicator Sequences

Pseudo-Long Memory Judgment of Per Capita Environmental Index Sequence

In the second stage, for achieving the more accurate estimates of the long memory or infinite memory parameters in BRICS, we control the deterministic trend and possible breakpoints of the ecological indicator sequences. **Table 4** provides the statistical test and long-memory parameter estimation results. The Mayoral test shown by **Table 4** identifies which hypothesis is true: non-stationary long memory vs. $I(0)$ + breakpoints for the environmental indicator series. The purpose of this test is to determine the cause resulting in the observed long-term memory processes: whether it is the result of a highly persistent shock, or whether those rare and unexpected events have changed the sequence's potential dynamics, inducing structural breakpoints. For each country, the first line reports the memory

parameter estimates, while the second and third lines report the critical values generated by the corresponding methods, i. e. Monte Carlo and Bootstrap approaches. We discuss presence or absence of structural breakpoints in the context of ARFIMA models, and are particularly interested in the effect of structural breakpoints on fractional integration parameter estimates. The results in **Table 4** show that the inference of whether or not per capita environmental indicator sequences are long memory is very robust. According to the statistics $R_b^f\left(\hat{d}_T\right)$ of per capita ecological footprints of BRICS, there is no evidence to reject the null hypothesis that the per capita EF sequences are fractionally integrated process (FI) with long memory characteristics in BRICS countries. Therefore, we can conclude safely that these fractionally integrated sequences might be non-covariance-stationary, with no significant breakpoints in the deterministic trend detected (as shown in **Figure 1**).

In China and India, with respect to the per capita BC sequences, we cannot reject the null hypothesis that the sequences of Equation (7) are long memory processes by the Mayoral test, depending Monte Carlo and bootstrap critical values. As shown in **Figure 2**, the per capita BCs are really long memory processes. For Brazil and South Africa, the Mayoral statistics is calculated by using the memory parameter estimates obtained from the detrended 2ELWd method. Both Monte Carlo and bootstrap critical values show that there is a breakpoint in per capita BCs, indicating that per capita BCs converge faster. Although the memory parameters estimated by the LW method in Russia show the existence of structural breakpoints, for the LW method is a relatively ineffective method, we generally prefer to the estimation results of the improved LW method. Therefore, it can be concluded that there is no breakpoint for the per capita BC in Russia. The economic implication of this conclusion is very clear: the Mayoral test indicates that there is no breakpoint in most countries' environmental indicators, which means that environmental indicators are highly persistent processes.

Therefore, on the whole, the overwhelming majority of BRICS countries have not undergone structural changes in the per capita environmental indicators, and they are long memory processes. Implications arising from the results above that international multilateral agreements such as the Kyoto Protocol and the Paris Agreement, which restrict greenhouse gas emissions and promote sustainable development, have not played a substantive role.

Pseudo-Long Memory Statistical Test of per Capita Environmental Indicator Sequences

In this paper we use inverting (single) bootstrap tests to construct three confidence intervals named by H0: Monte Carlo, H0: Bootstrap and H1: Bootstrap (see notations shown by legends of **Figures 5**, **6**) for the Mayoral test statistics in 5 BRICS countries and compare with each other and infer whether the H0 hypothesis hold at usual significant levels. The results presented in **Figures 5**, **6** suggest that CIs are constructed more profitably under H0 than under H1.

Of the three confidence intervals, the evaluation of H0: Monte Carlo and H0: Bootstrap is based on their coverage accuracy and average length. Column 2 and 3 of **Figures 5** and **6** show average lengths and the coverage probabilities of the confidence intervals calculated for 2,000 runs, respectively. A shorter average length corresponds to a better effectiveness performance. **Figure 5** shows that the Monte Carlo test statistics produces CIs that are longer (lower effectiveness) than CIs constructed using the bootstrap test statistics for EF indicator in China, India and Brazil; while MC critical values are of minor advantages over Bootstrap counterparts in South Africa and Russian based on effective curves.

The results set out in **Figure 6** show that the confidence interval of the Mayoral test statistics built by Monte Carlo simulation has the better performance on coverage probability compared with that of the bootstrap method for Mayoral test in South Africa and Russia. Just the opposite is true to other BRICS countries. However, the results displayed in **Figure 6** reveal that for the BC indicators' Mayoral test statistics, the average length of the Monte Carlo confidence intervals is longer than that produced by the Bootstrap method in Brazil, indicating worse effectiveness. The inference dependent of Bootstrap critical values is more effective in terms of higher coverage probability and shorter average length than Monte Carlo counterparts of other countries among BRICS. On the contrary, the average length of the Monte Carlo confidence intervals is shorter in South Africa, which means Monte Carlo critical values are more satisfactory inference tools. The effectiveness curves do not show the visually obvious difference in other three BRICS countries. Synthesizing the results of **Figures 5, 6**, and balancing the performance of the effectiveness and coverage probability, Monte Carlo critical values are appropriately able to be used in Mayoral test in South Africa and Russia. Thus, the exact bootstrap method is not recommended for identifying potential break points in the two countries. By contrast, bootstrap critical values are the preferred inference tools to Mayoral test in remaining countries. No matter what environmental indicators are tested, these conclusions from **Figures 5**, **6** are the same.

Breakpoint Identification of per Capita Environmental Indicator Sequences (Berkes Test)

Breakpoint Judgment of per Capita Environmental Indicator Sequences

In order to check the reliability of the results, this paper supplements the Mayoral (2) test by using the CUSUM statistic of Berkes et al. (1) to perform a further test so as to infer the existence of breakpoints and determine the locations of breakpoints. **Table 5** shows the statistics M_n of Berkes et al. (1). Due to the good performance of small samples, we use the Bartlett kernel function and select the optimal bandwidth suggested by Newey-West (see Section **Convergence Pseudo-long Memory Test of Relative Environmental Indicators Per Capita**). The null hypothesis represented by Equation (10) indicates that there is no structural change in the deterministic trend of the ecological indicator sequence. **Figure 7** shows how the ecological footprint statistics M_n for each country evolves over time. Due to the space limitation, the statistics M_n of the BCs per capita in BRICS

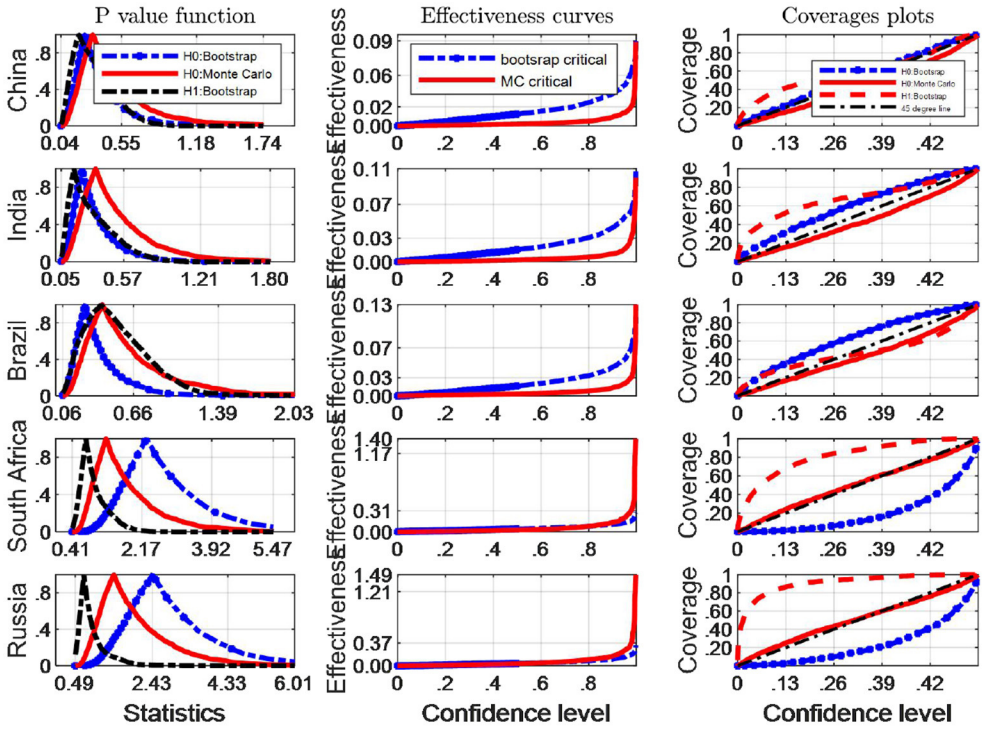


FIGURE 5 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Mayoral test using EF indicators.

countries over the sample years is not listed in the present paper. The statistics in **Figure 7** and **Table 5** shows that the maximum value of the Berkes statistics of the per capita EF sequences and per capita BC sequences of the BRICS countries will not be lower than critical values of themselves, which can reject the null hypothesis and there are not any break points identified. Therefore, within the sample region of 57 years, the ecological footprints per capita and BC sequences per capita are either covariance non-stationary process of mean reverting or unit root process, i.e., there is long memory.

Breakpoint Statistical Test of per Capita Environmental Indicator Sequences

Figures 8, **9** show coverage and effectiveness results for EF and BC indicators respectively, with separate plots for Monte Carlo and Bootstrap critical values. These two figures show that the Monte Carlo arbitrament methods tended to provide smaller confidence intervals with better coverage accuracy than the bootstrap arbitrament methods over almost all Berkes test statistics domains (judged by the first column and the third column of **Figure 8**).

The two figures again indicate that Monte Carlo method coverage probability generally compare favorably against those of Bootstrap method based on inverting Bootstrap CIs, and in particular Monte Carlo method intervals seem to emerge with slightly better coverage rates asymptotically from their evolving dynamics as well. However, the Monte Carlo method indiscernibly performs better than does the Bootstrap methods when the Berkes test methodology is employed to the Russian EF indicator and the South African BC indicator. These plots also indicate that, comparing with the Bootstrap method, although the Monte Carlo inference method has similar coverage rates over higher confidence levels, there are obvious systematic differences in overall interval lengths. Over lower confidence levels, we can identify similar effectiveness of the two inference methods generally.

When we test potential breakpoints occurring in BC indicator sequences, in terms of interval lengths, the Monte Carlo method has uniformly narrower intervals, especially when confidence levels are small (**Figure 9**). Following the results provided by **Figure 8**, the Monte Carlo critical values has the most satisfactory "true" effectiveness for the inference of potential break points in almost all the BRICS countries, expect for South Africa and Russia when EF indicators are tested. However, the difference between the average lengths of the various methods is not very large, especially over lower confidence levels. Consequently, the Monte Carlo critical values should be retained to judge whether a long-memory process is true or just a mean-shifted process, on the double basis of the coverage and the average length criteria. With respect to the BC indicator, the confidence intervals for critical values from the Monte Carlo methods are shorter than the bootstrap method at usual confidence levels, e.g., 95%.

Our simulation results indicate that when the BC indicators are tested if they are pseudo-long memory processes, the Monte Carlo critical values have the best performance based on both coverage probability compared with the bootstrap

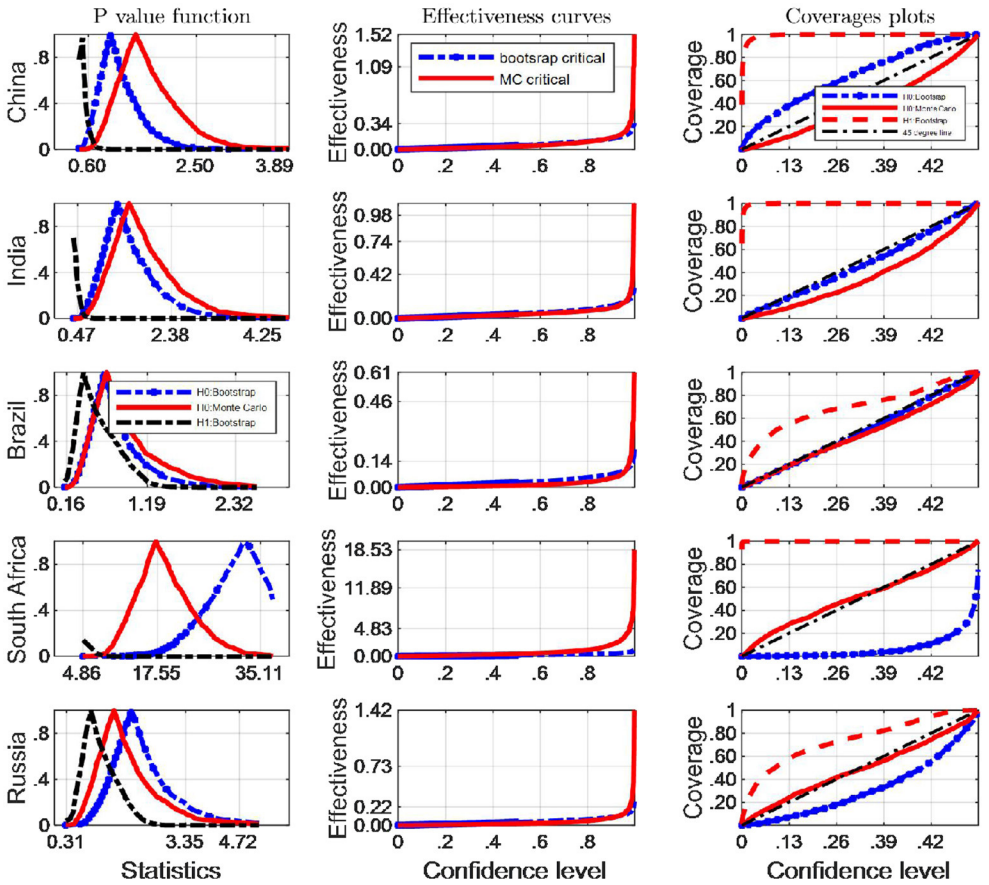


FIGURE 6 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Mayoral test using BC indicators.

counterparts, since the coverage plots produced by MC critical values are closer to 45° lines. However, what the second column tells is that more statistory effectiveness curves are produced by bootstrap critical values expect for India and Brazil. A summary statement can be concluded that Monte Carlo CIs generally performed better than their bootstrap counterparts. Based on the analysis above, when the EF indicators are tested pseudo-long memory, the bootstrap critical values should be resorted to in South Africa and Russia, and the Monte Carlo critical values are appropriate to other BRICS countries. Therefore, this study recommends Monte Carlo confidence interval to quantify the uncertainty of the Berkes statistics, and more robust inferences based Monte Carlo critical values.

Combined with the Mayoral and Berkes pseudo-long memory tests of the environmental indicators (EF and BC) of each BRICS country, the results provide consistent evidence that only China and Russia have breakpoints in their per capita EF (the China's per capita EF breakpoint occurred in 1990; The per capita EF breakpoint in Russia occurred in 1993). Therefore, a large number of countries follow the evolutionary trend of long memory processes. These breakpoints correspond to the late 1970s and early 1980s. During this period, the Iranian revolution and Iran-Iraq war broke out in 1978–1979. These two major oil producers, accounting for 6% of the world's oil output, experienced a decline in oil products. The real price of oil doubled between 1978 and 1981, which had a negative impact on the economies of BRICS countries where the ecological indicator breakpoints occurred. In response to the oil price shock, BRICS countries' ecological footprints fell significantly, following their economic recessions. We can also see that the two significant structural breakpoints in China correspond to the 1990s, when events, including the Gulf War and coming into force of the Kyoto Protocol at the end of 1997, affected the ecological indicators. However, the break point in the rising process of Russian per capita EF sequence is due to taking the former Soviet Union data as a proxy before 1991, which resulted in sudden shifts in per capita ecological indicators between 1992 and 1995.

Tables 2, 3 summarize the results judged by the estimated values of *d*. The results show the ecological indicators are long memory in all BRICS countries, and the subjects we selected are good for this study. The BRICS are grouped into three categories in term of the degree of long memory. The first group refers to the countries belonging to the

TABLE 5 | Pseudo long memory Berkes test of relative per capita environmental indicators in BRICS.

	LW		ELW		De-trented (breakpoint) 2ELWd	
	EF	BC	EF	BC	EF	BC
China						
Berkes Statistics	4.712	3.212	4.713	3.612	2.13	3.812
MC CV	0.665	1.732	0.668	1.730	2.665	1.729
Bootstrap CV	1.165	1.868	1.166	1.864	2.161	1.867
India						
Berkes Statistics	4.105	4.104	4.105	4.108	4.105	4.108
MC CV	1.055	1.857	1.048	1.858	1.038	1.858
Bootstrap CV	1.891	1.925	1.873	1.924	1.913	1.924
Brazil						
Berkes Statistics	4.375	4.105	4.375	4.105	4.374	4.105
MC CV	0.905	2.243	0.903	2.242	0.902	2.240
Bootstrap CV	1.333	2.291	1.340	2.283	1.350	2.286
South Africa						
Berkes Statistics	2.735	4.102	2.735	4.102	2.735	2.105
MC CV	0.823	2.616	0.824	2.616	0.822	2.616
Bootstrap CV	1.177	2.561	1.172	2.562	1.179	2.562
Russia						
Berkes Statistics	4.117	4.111	4.117	4.111	2.917	2.214
MC CV	0.849	1.481	0.848	1.479	3.251 (–)	2.491 (–)
Bootstrap CV	1.493	1.667	1.517	1.685	3.132 (–)	2.678 (–)

'mean reversion' case (i.e., South Africa and Russia in term of CFs, and China, India and South Africa in term of BCs with an order of integration significantly smaller than 1); the second group refers to those countries showing evidence of unit roots (i.e., Brazil and Russia in term of BCs with orders of integration around 1); and the third group refers to countries with divergent ecological indicators (China, India and Brazil in term of CFs with orders of integration significantly above 1).

It is not surprising that the per capita ecological footprints of the vast majority of BRICS countries reported in the empirical results are non-stationary and non-mean-reverting. This reflects that the per capita ecological footprints develop along a trend path, usually along an upward trajectory, and does not return to their means during the sample period. However, due to heavily consuming natural resources, in BRICS countries the ecological carrying capacities evolve along the downward trend. This indicates that these countries have continuously imposed more pressures on the per capita ecological footprints and ecological carrying capacities, resulting in the ecological footprints exceeding the biological carrying capacities of these countries. Therefore these countries uniformly fall into an ecological deficit state, as shown in **Figure 2**. Some of these countries have high per capita ecological carrying capacities, but have a higher marginal consumption propensity when they are utilizing the nature. For example, Brazil has the highest per capita ecological carrying capacity, but will still faces the dilemma of an ecological deficit due to excessive demand for natural resources.

CONCLUSIONS AND IMPLICATIONS

The importance of a composite indicator of environmental degradation (ecological footprint) is that it can go beyond exparte discussions of climate change and sustainability, because it can also track demand for services provided by a wide range of natural resources and ecosystems, rather than just focus on the atmospheric carbon accumulation. In the contemporary world, people have used 1.6 earths to produce the resources consumed and dump the wastes discharged. This trend means that it will take the earth 1 year and 6 months to regenerate natural resources used and absorb wastes per year because of human activities. If this kind of consumption pattern is as of old, the total demand for ecosystems is expected to exceed 80% of natural carrying capacities in another 10 years (39).

Environmental economics is paying more and more attention to the convergence of environmental indicators such as ecological footprints, especially the topics related to natural deterioration. There are two reasons why convergence of environmental indicators is important. (1) The need for close cooperation between countries and regions with similar productivity characteristics when formulating and implementing environmental policies. In order to mitigate the threat of global warming, policies should not be limited to the territorial scope of just one country, but should take actions at the international level. The Tokyo Protocol and the Paris Agreement are examples of the international communities' joint resistance to environmental degradation. (2) The consumption patterns, foreign trade and industrial structure of countries or regions

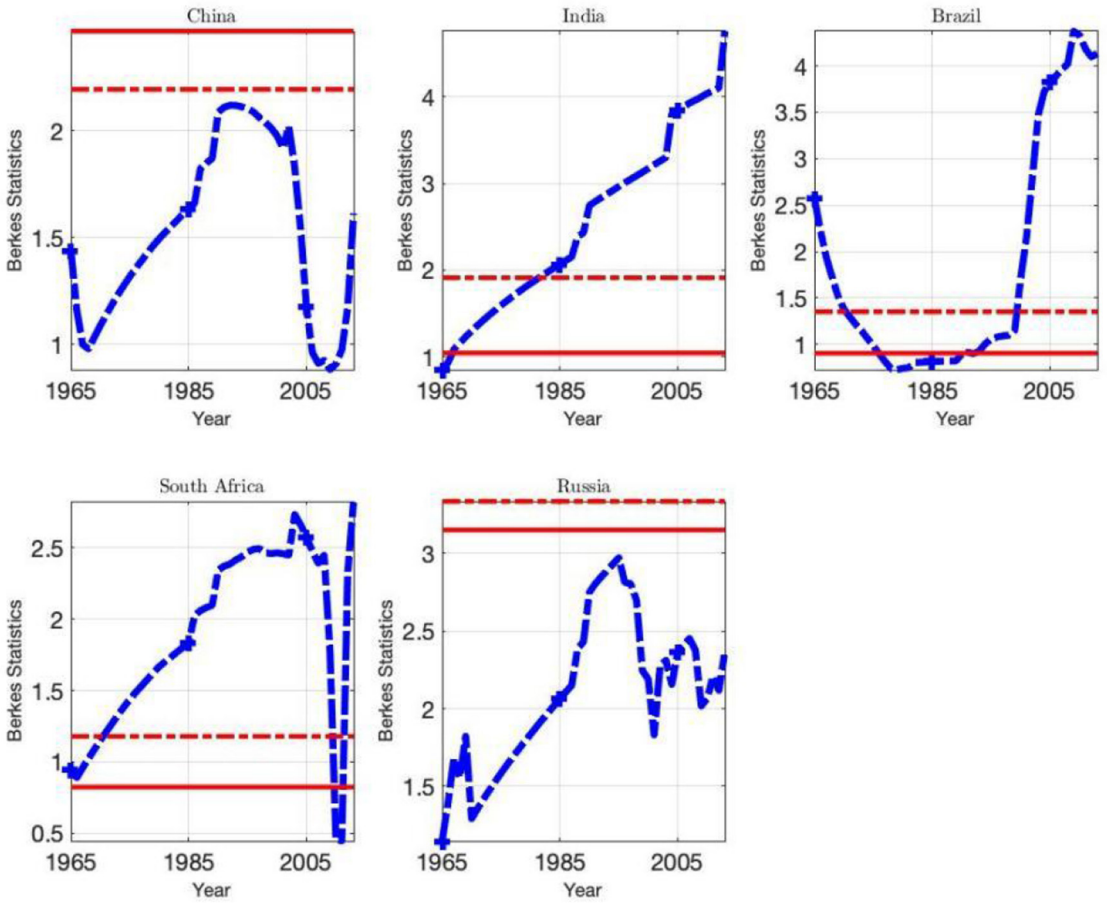


FIGURE 7 | M_n statistics for BRICS. This graph plots Berkes statistics (blue lines) evolving over time. Red solid lines represent Monte Carlo critical values at 5% significant level; while red dash lines represent bootstrap critical values at 5% significant level. When red lines are located above the maxima of *M* statistics, Berkes test reject null hypothesis. With respect to Berkes statistics represented by Equation (22), empirical critical values are obtained by 2,000 repetitions of Monte Carlo simulation and 2,000 repetitions Bootstrap resample respectively.

affect the evolving trend of environmental and ecological indicators. Therefore, it is urgent to identify the factors that determine the convergence or divergence of ecological indicators. Environmental policies to protect the nature have an obligation to change the deteriorating trend of environmental indicators. Therefore, identifying the mechanisms that trace the trajectories of changes in ecological indicators is a topic of interest to both policymakers and academic researchers.

This present paper applies the sample data from 1961 to 2017 to study whether or not the stochastic convergence occurred in the ecological footprint and ecological carrying capacity series of BRICS countries. Two rigorous statistical testing methods are employed to identify whether or not the long memory processes (if exist) characterizing ecological indicators are falsely modeled (i.e., identified break points). Through various frequency-domain Whittle estimators and statistical tests, it is found that the fractional integration parameters of the vast majority of per capita EF sequences are greater than 1, indicating that these environmental indicators are non-mean reverting (divergence); The parameters of fractional integration of BC sequences are mostly within the interval $0 < d \leq 1$, which indicates that the ecological footprints are non-stationary long-memory processes. The fractional integration parameters of ecological carrying capacities are obviously smaller than those of ecological footprints, but they are still in the non-stationary and long memory domain. The significant discrepancy to the EF indicators in most countries, however, is that the BC sequences are convergent processes, namely mean-reverting. By comparisons, it is found that the convergence of ecological footprints and ecological carrying capacities is closely related to the economic development of various countries. Generally speaking, richer countries have a higher convergence rate of these two indicators being discussed; As the world's largest developing countries, the BRICS countries' ecological indicators are mostly divergent.

From the perspective of existing or potential policies that prevent ecological indicators worsening, the conclusions of this study are very informative. For most BRICS countries, the ecological footprints exhibit non-stationary behaviors, which

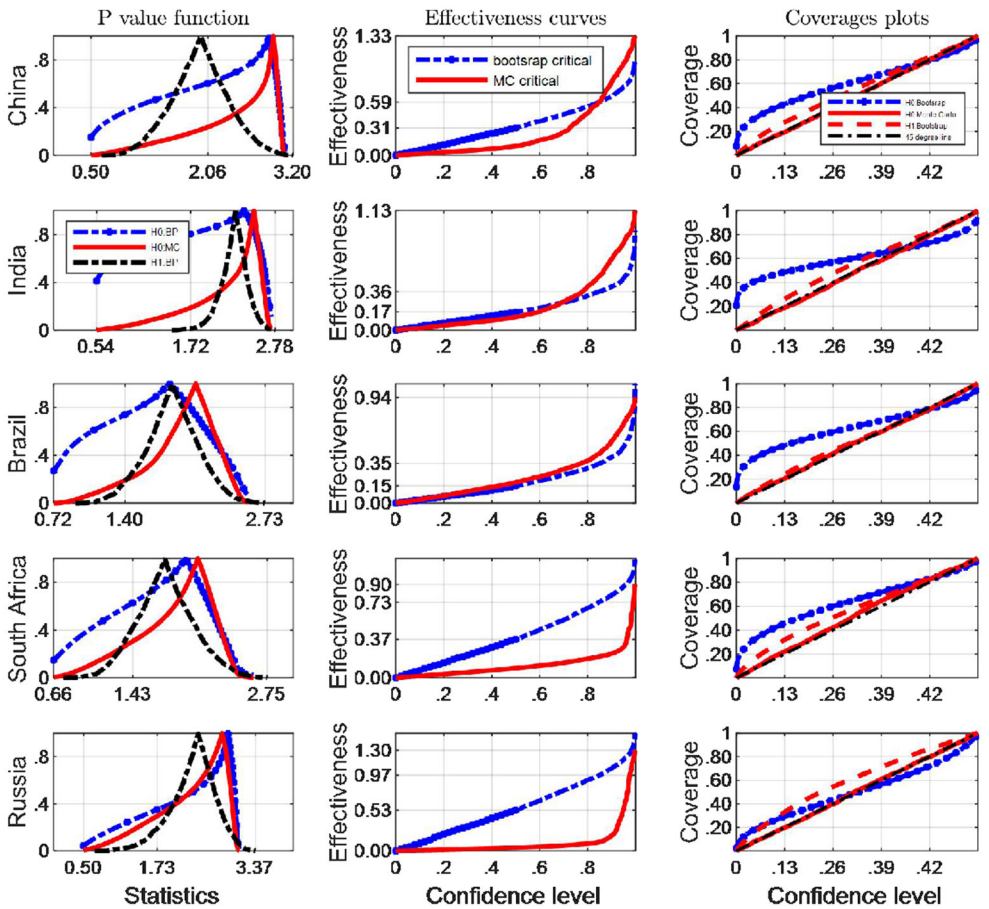


FIGURE 8 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Berkes test using EF indicators.

fully demonstrates that policies that affect the ecological footprints actually have long-term and permanent effects. There are several policies that can address the growing ecological footprints. Such a series of policies includes a carbon tax designed to reduce the demand for carbon and governmental subsidies expended to guarantee the consumption of alternative clean energies. Controls on land use should also be strengthened to protect land, forests and other natural habitats. Regardless of whether or not certain environmental policies are temporary in nature, the fact that the ecological footprints are non-stationary ensures governmental policies can take long-term effects on these indicator's dynamics. China's per capita EF series have undergone structural changes during the sample period, indicating that the rising rate is faster than other BRICS countries. Implications arising from this observation a more stringent and long-term policy tool package should be adopted for energy conservation and carbon emission reduction. The per capita BC sequence in South Africa has also undergone structural changes in the process of persistent decline (judged by the Mayoral test), anticipating a faster decreasing rate. Therefore, in the pursue of economic development, Brazil should get rid of over-exploitating nature and adopt stricter environmental protection policies to prevent the further decline of its ecological carrying capacity.

Because the ecological footprints are found to be non-stationary in most BRICS countries, predicting their evolving dynamics based on past behavior may constitute a great obstacle (if not be feasible). However, their non-stationary behaviors means that policies aimed at changing it will have lasting and persistent impacts. As the vast majority of macroeconomic variables are also non-stationary, especially for level values (not differenced values), the ecological footprints may have the same fractional integration order as that of macroeconomic variables, thus indicating the existence of cointegration between these variables, in other words, long-term equilibrium relationship, which can assist in formulating relevant environmental policies.

Convergence of environmental indicators between countries is considered as a fair share of carbon dioxide emission rights. However, different countries have different natural endowments, and furthermore resources can be transferred between different countries. The stationarity of a sequence provides information about whether the effects of a shock to the sequence is temporary or permanent. A stationary sequence manifest itself that the effects of this shock are temporary. Therefore, if the impacts of

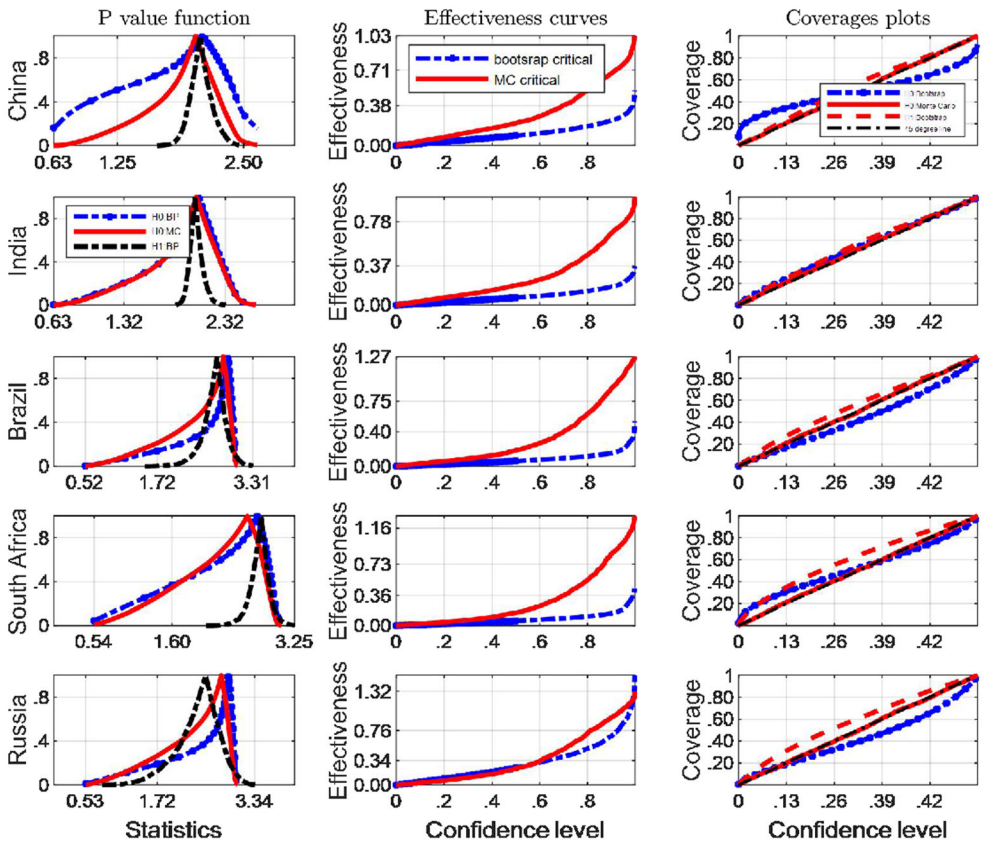


FIGURE 9 | A comparison of confidence intervals of Monte Carlo and counterparts of bootstrap for the Berkes test using BC indicators.

shocks to the ecological footprint sequences dissipate quickly, these sequences of various countries converge to the sample means stochastically. A non-stationary sequence indicates that the effects of the shock hitting it are permanent, i.e. this sequence will deviate from the sample mean eventually. Similarly, when a ecological indicator is non-stationary, environmental policies that affect it have persistent, and even permanent effects. At the same time, the convergence of ecological indicators between countries shows that it is necessary to adopt internationally cooperative policies to protect the vulnerable environment. The lack of convergence indicates that these policies are out of permanent effects and the cross-broad policies are not very powerful.

Due to the long-memory characteristics of environmental indicators, the impacts of a positive shock (usually comes from an environmental policy) will be long-term and this policy effect will be better. For example, specific policies include energy efficiency improvement projects and subsidies for renewable energies. However, the divergence of environmental indicators is more disadvantageous. If an adverse random impact occurs, the impact will be very long-term, and a stronger policy intervention is needed to reverse the trajectory of environmental deterioration. In order to deter the worsening trend of ecological indicators, a steadfast policy stance is necessary. Therefore, we argue that: (1) A policy that is designed to be temporary or eventually revoked cannot expect to have lasting effects. (2) A policy that is either permanent or believed to be permanent will often have stronger and more persistent effects. Based on our empirical results, the degree of commitment is as important or more important than the validity or the potential effects of a policy.

In addition, some country-specific characteristics, such as the degree of economic development, industrial structure and consumption patterns, should also be taken into account. As high-income countries and developed countries formulate some better policies and more effective measures to confront the challenge of the ecological indicator worsening, low-income countries should borrow some advanced experiences of developed countries who share similar evolving trajectories of ecological indicator to achieve greater progress in curbing the deterioration of ecological indicators.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary

material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

We acknowledge financial support from the research project A Research on Parameter Identification and Application of DSGE Models Allowing for Indeterminacy (Approval Number: 71863008) sponsored by National Natural Science Foundation of China.

REFERENCES

1. Berkes I, Horváth L, Kokoszka P, Shao Q. On discriminating between long-range dependence and changes in mean. *Ann Stat.* (2006) 34:1140–65. doi: 10.1214/009053606000000254
2. Mayoral L. Testing for fractional integration versus short memory with structural breaks. *Oxford Bull Econ Stat.* (2012) 74:278–305. doi: 10.1111/j.1468-0084.2011.00645.x
3. Grossman GM, Krueger AB. *Environmental Impacts of North American Free Trade Agreement.* (1991). NBER Working Paper Series No.: 3914.
4. Brock WA, Taylor MS. *The Kindergarten Rule of sustainable growth.* NBER, Massachusetts (Working Paper 9597) (2003).
5. Robinson PM. Efficient tests of nonstationary hypotheses. *J Am Statis Assoc.* (1994) 89:1420. doi: 10.2307/2291004
6. World Bank (2019). *World Development Indicators.* Available online at: https://databank.worldbank.org/source/world-development-indicators
7. Wang Z. Does biomass energy consumption help to control environmental pollution? Evidence from BRICS countries. *Sci Total Environ.* (2019) 670:1075–83. doi: 10.1016/j.scitotenv.2019.03.268
8. Ulucak R, Lin D. Persistence of policy shocks to ecological footprint of the USA. *Ecol Indic.* (2017) 80:337–43. doi: 10.1016/j.ecolind.2017.05.020
9. Ulucak R, Apergis N. Does convergence really matter for the environment? An application based on club convergence and on the ecological footprint concept for the EU countries. *Environ Sci Policy.* (2018) 80:21–7. doi: 10.1016/j.envsci.2017.11.002
10. Bilgili F, Ulucak R. Is there deterministic, stochastic, and/or club convergence in ecological footprint indicator among G20 countries? *Environ Sci Pollut Res.* (2018) 25:35404–19. doi: 10.1007/s11356-018-3457-1
11. Solarin SA. Convergence in CO_2 emissions, carbon footprint and ecological footprint: evidence from OECD countries. *Environ Sci Pollut Res.* (2019) 26:6167–81. doi: 10.1007/s11356-018-3993-8
12. Yilanci V, Gorus MS, Aydin M. Are shocks to ecological footprint in OECD countries permanent or temporary? *J Clean Prod.* (2019) 212:270–301. doi: 10.1016/j.jclepro.2018.11.299
13. Yildirim DÇ, Yildirim S, Erdogan S, Demirtas I, Couto G, Castanho RA. Time-varying convergences of environmental footprint levels between European Countries. *Energies.* (2021) 14:1813. doi: 10.3390/en14071813
14. Solarin SA, Bello MO. Persistence of policy shocks to an environmental degradation index: the case of ecological footprint in 128 developed and developing countries. *Ecol Indic.* (2018) 89:35–44. doi: 10.1016/j.ecolind.2018.01.064
15. Bilgili F, Ulucak R, Koçak E. Implications of environmental convergence: continental evidence based on ecological footprint. In: *Energy and Environmental Strategies in the Era of Globalization.* Cham: Springer (2019). p. 133–65.
16. Phillips PC, Sul D. Transition modeling and econometric convergence tests. *Econometrica.* (2007) 75:1771–855. doi: 10.1111/j.1468-0262.2007.00811.x
17. Solarin SA, Tiwari AK, Bello MO. A multicountry convergence analysis of ecological footprint and its components. *Sustain Cities Soc.* (2019) 46:101422. doi: 10.1016/j.scs.2019.101422
18. Ozcan B, Ulucak R, Dogan E. Analyzing long lasting effects of environmental policies: evidence from low, middle and high income economies. *Sustainable Cities and Society.* (2019) 44:130–43. doi: 10.1016/j.scs.2018.09.025
19. Erdogan S, Okumus I. Stochastic and club convergence of ecological footprint: an empirical analysis for different income group of countries. *Ecol Indic.* (2021) 121:107123. doi: 10.1016/j.ecolind.2020.107123
20. Işik C, Ahmad M, Ongan S, Ozdemir D, Irfan M, Alvarado R. Convergence analysis of the ecological footprint: theory and empirical evidence from the USMCA countries. *Environ Sci Pollut Res.* (2021) 2021:1–12. doi: 10.1007/s11356-021-12993-9
21. Yilanci V, Pata U. Convergence of per capita ecological footprint among the ASEAN-5 countries: evidence from a non-linear panel unit root test. *Ecol Indic.* (2020) 113:106178. doi: 10.1016/j.ecolind.2020.106178
22. Tillaguango B, Alvarado R, Dagar V, Murshed M, Pinzón Y, Méndez P. Convergence of the ecological footprint in Latin America: the role of the productive structure. *Environ Sci Pollut Res.* (2021) 28:59771–83. doi: 10.1007/s11356-021-14745-1
23. Barassi MR, Spagnolo N, Zhao Y. Fractional integration versus structural change: testing the convergence of CO_2 emissions. *Environ Resour Econ.* (2017) 71:923–68. doi: 10.1007/s10640-017-0190-z
24. Lin D, Hanscom L, Martindill J, Borucke M, Cohen L, Galli A, et al. *Working Guidebook to the National Footprint Accounts*. Oakland: Global Footprint Network (2018).
25. Robinson PM. Gaussian semiparametric estimation of long range dependence. *Ann Stat.* (1995) 23:1630–61. doi: 10.1214/aos/1176324317
26. Phillips PCB, Shimotsu K. Local Whittle estimation in nonstationary and unit root cases. *Ann Stat.* (2004) 32:656–92. doi: 10.1214/009053604000000139
27. Shimotsu K, Phillips PCB. Exact local Whittle estimation of fractional integration. *Ann Stat.* (2005) 33:1890–933. doi: 10.1214/009053605000000309
28. Perron P. The great crash, the oil price shock, and the unit root hypothesis. *Econom J Econ Soc.* (1989) 58:1361–401. doi: 10.2307/1913712
29. Leybourne SJ, Newbold P. Spurious rejections by cointegration tests induced by structural breaks. *Appl Econom.* (2003) 35:1117–21. doi: 10.1080/0203684032000082068
30. Granger C, Terasvirta T. A simple nonlinear time series models with misleading linear properties. *Econ Lett.* (1999) 62:161–5. doi: 10.1016/S0165-1765 (98)00228-6
31. Charfeddine L, Guégan D. Breaks or long memory behaviour: an empirical investigation. *Phys A, Stat Mech Appl.* (2012) 391:5712–26. doi: 10.1016/j.physa.2012.06.036
32. Strazicich MC, List JA. Are CO_2 emission levels converging among industrial countries? *Environ Resour Econ.* (2003) 24:263–71. doi: 10.1023/A:1022910701857
33. Robinson PM. Effificient tests of nonstationary hypotheses. *J Am Stat Assoc.* (1994) 89:1420–37. doi: 10.1080/01621459.1994.10476881
34. Andrews DWK. Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica.* (1991) 59:817–58. doi: 10.2307/2938229
35. Newey WK, West KD. Automatic lag selection in covariance matrix estimation. *Rev Econ Stud.* (1994) 61:631–53.
36. Aue A, Horváth L, Hušková M, Ling S. On distinguishing between random walk and change in the mean alternatives. *Econ Theory.* (2009) 25:411–41. doi: 10.1017/S0266466608090130
37. Wackernagel M, Zokai G, Iha K, Kelly R, Ortego J. *The Footprint and Biocapacity Accounting: Methodology Background for State of the States 2015.* Oakland: Global Footprint Network (2015).
38. Granger CWJ, Joyeux R. An introduction to long-memory time series models and fractional differencing. *J Time Ser Anal.* (1980) 1:15–29. doi: 10.1111/j.1467-9892.1980.tb00297.x

39. National Footprint Accounts Working Guidebook to the National Footprint Accounts. (2017). Available online at: http://www.footprintnetwork.org/content/documents/National_Footprint_Accounts_2017_Guidebook.pdf2017

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fpubh.2022.887123

Application of Smart Healthcare in Comparative Analysis of Effect of Early External Fixator and Plate Internal Fixation Treatment on Postoperative Complications and Lower Limb Function Recovery of Patients With Unstable Pelvic Fracture

Hongwei Zhang[1†], *Fang Sun*[2†] *and Yao Li*[3*]

[1] Department of Emergency, The First Affiliated Hospital of Jiamusi University, Jiamusi, China, [2] Plastic Burn Beauty Center, The First Affiliated Hospital of Jiamusi University, Jiamusi, China, [3] Department of Orthopeadic Surgery, The First Affiliated Hospital of Jiamusi University, Jiamusi, China

Objective: To comparatively analyze the effect of early external fixator and plate internal fixation treatment on postoperative complications and lower limb function recovery of patients with unstable pelvic fracture based on smart healthcare.

Methods: The clinical data of 98 patients with unstable pelvic fractures treated in our hospital from August 2018 to August 2021 were collected for retrospective analysis, and the patients were split into group A (plate internal fixation, $n = 49$) and group B (early external fixator treatment, $n = 49$) according to the surgical modalities. The time of operation, intraoperative blood loss, postoperative complications, etc. were compared between the two groups.

Results: Between the two groups, group B had significantly better clinical indicators ($P < 0.001$), a greatly higher good rate of fracture reduction and postoperative Harris score ($P < 0.05$), and obviously lower VAS score and total incidence rate of postoperative complications ($P < 0.05$).

Conclusion: Through the analysis based on smart healthcare, it is found that compared with plate internal fixation treatment, early external fixator treatment presents a better effect in treating patients with unstable pelvic fracture, because it is a reliable method to accelerate fracture healing, reduce postoperative complications, and improve lower limb function.

Keywords: early external fixator, plate internal fixation, unstable pelvic fracture, postoperative complications, lower limb function recovery, smart healthcare

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Yejun Liu,
Qianjiang Central Hospital of Hubei Province, China
Kadiyala Ramana,
Annamacharya Institute of Technology and Sciences, India

***Correspondence:**
Yao Li
liyao@jmsuf.com.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *01 March 2022*
Accepted: *21 March 2022*
Published: *29 April 2022*

Citation:
Zhang H, Sun F and Li Y (2022) Application of Smart Healthcare in Comparative Analysis of Effect of Early External Fixator and Plate Internal Fixation Treatment on Postoperative Complications and Lower Limb Function Recovery of Patients With Unstable Pelvic Fracture. Front. Public Health 10:887123. doi: 10.3389/fpubh.2022.887123

INTRODUCTION

Pelvic fractures account for approximately 3% of all fractures. Unstable pelvic fractures are classified as high-energy injuries (1) and are often associated with other complications, with high mortality and disability rates. Conservative treatment is often selected in clinics to treat unstable pelvic fractures, but the incidence rates of complications such as hypostatic pneumonia and gastrointestinal hemorrhagic stress ulcer are higher (2), which, despite current improvements in medical technology, still affect the prognosis of patients, resulting in a severe reduction in their quality of life (QOL) (3, 4). Since the pelvic bone is not stable and relies mainly on soft tissues such as ligaments, so the key to treatment is to provide it with an environment that is mechanically stable throughout its structure, thus aiding the recovery of both the bone and soft tissues (5). With the progress of Internet technology, China proposes to develop the smart health industry and promote the deep integration of cloud computing, big data, mobile Internet, and the health services industry, under which context smart healthcare was born. Smart healthcare enables a more scientific management model of daily life and production that greatly improves the distribution and utilization efficiency of resources. Plate internal fixation is currently a common method for the clinical treatment of unstable pelvic fractures, and clinical studies (6) have found that although it has a better fixation effect, there are disadvantages such as a long time of surgery, high difficulty in surgical manipulation, long time of incision healing and more postoperative complications. The external fixation technique is mostly used in the first aid of pelvic fractures because it plays an important role in the stability and hemodynamics of pelvic fractures and saves valuable time for the management of combined injuries, and its effect has been demonstrated in some fracture types, such as femur fractures, tibial shaft fractures, and comminuted intra-articular fractures of the distal radius in children (7–9). Currently, there are fewer clinical studies on the external fixation treatment for pelvic fractures. The study, based on smart healthcare, further summarized and concluded the diagnosis and surgical treatment for fracture types to better apply the external fixation technique in pelvic fractures, and meanwhile, measured the best position and angle for external fixator screw installation according to patients' three-dimensional CT images of the pelvis before surgery, in the hope of providing greater theoretical and data support for the surgery.

MATERIALS AND METHODS

General Data

The study objects were patients with unstable pelvic fracture treated in our hospital. All patients received relevant examinations after admission, see **Figure 1** for the technical route. The study met the *World Medical Association Declaration of Helsinki (2013)* (10).

Inclusion and Exclusion Criteria

Inclusion criteria. ① The images met the X-ray signs of unstable pelvic fracture; ② the time from injury to surgery $\leq$ 3 weeks; ③ the patients met the indications of surgical treatment; and ④ the patients had complete clinical data.

Exclusion criteria for the patients. ① Pathological or old fracture; ② complicated with severe head trauma, organic injury, or neurovascular injury; and ③ severe osteoporosis.

Therapies

Patients in group A received internal fixation with steel plate, and the steps were as follows. General anesthesia was performed, before surgery, 0.08 mg/kg of midazolam (manufacturer: Yichang Humanwell Pharmaceutical Co., Ltd.; NMPA approval no. H20067040; specification: 2 ml: 2 mg) was administered via intramuscular injection, and 0.03 mg/kg of midazolam, 3 μg/kg of fentanyl (manufacturer: Jiangsu Nhwa Pharmaceutical Co., Ltd.; NMPA approval no. H20113509; specification: 10 ml: 0.5 mg), 1.5 mg/kg of propofol (manufacturer: Sichuan Guorui Pharmaceutical Co., Ltd.; NMPA approval no. H20040079; specification: 10 ml: 0.1 g) and 0.6 mg/kg of rocuronium (manufacturer: Guangdong Jiabo Pharmaceutical Co., Ltd.; NMPA approval no. H20183109; specification: 5 ml: 50 mg) was given *via* an intravenous drip. For patients with pure anterior ring injury such as separation of the pubic symphysis, the Pfannenstiel incision approach was adopted to expose the pubis and pubic symphysis site, the tissues inside the inguinal canal were well protected, and pelvic reduction was performed with pelvic reduction forceps or screw reduction forceps, the pelvic reconstructive plate was used for fixation, which was placed above the pubic symphysis site. For patients with fractures at the root of the superior branch of the pubis and acetabulum, an ilioinguinal approach was adopted, the femoral veins, femoral arteries, and peripheral nerves were well protected, after reduction, fixation was performed with the reconstructive plate, which was placed above the pubic branch, and C-arm X-ray was used to confirm that the screws were not in the acetabulum. For patients with pubic branch fracture at the anterior pelvic ring or mild separation of the pubic symphysis, if the pelvis was stable after reduction and fixation of the posterior ring, the method of single posterior ring fixation could be adopted (11). For patients with subluxation or luxation of the sacroiliac joint, an ilioinguinal approach was adopted to expose the front of the whole sacroiliac joint and iliac fossa, subperiosteal dissection was performed, the Kirschner wire (manufacturer: Tianjin Jinxingda Industries Co., Ltd.; model: type ZQY) was screwed in sacrum from the outside of L_5 nerve root, and fixation was conducted with 2- to 4-hole steel plate. For patients with sacroiliac joint luxation or sacrum fracture, a posterior exposure approach was performed, the patients were in the supine position, a straight or arc-shaped incision was made along the sacroiliac joint, the gluteus maximus was separated from the attachment point, and the superior gluteal blood vessels and nerves were protected. Closed reduction was mostly adopted for the fixation of iliosacral screws, the patients were in the supine position or prone position, the C-arm system confirmed that the position, direction, and length were appropriate in 3 planes on the lateral part, pelvic inlet, and outlet, and the corresponding screws were selected for fixation of fracture and luxation (12).

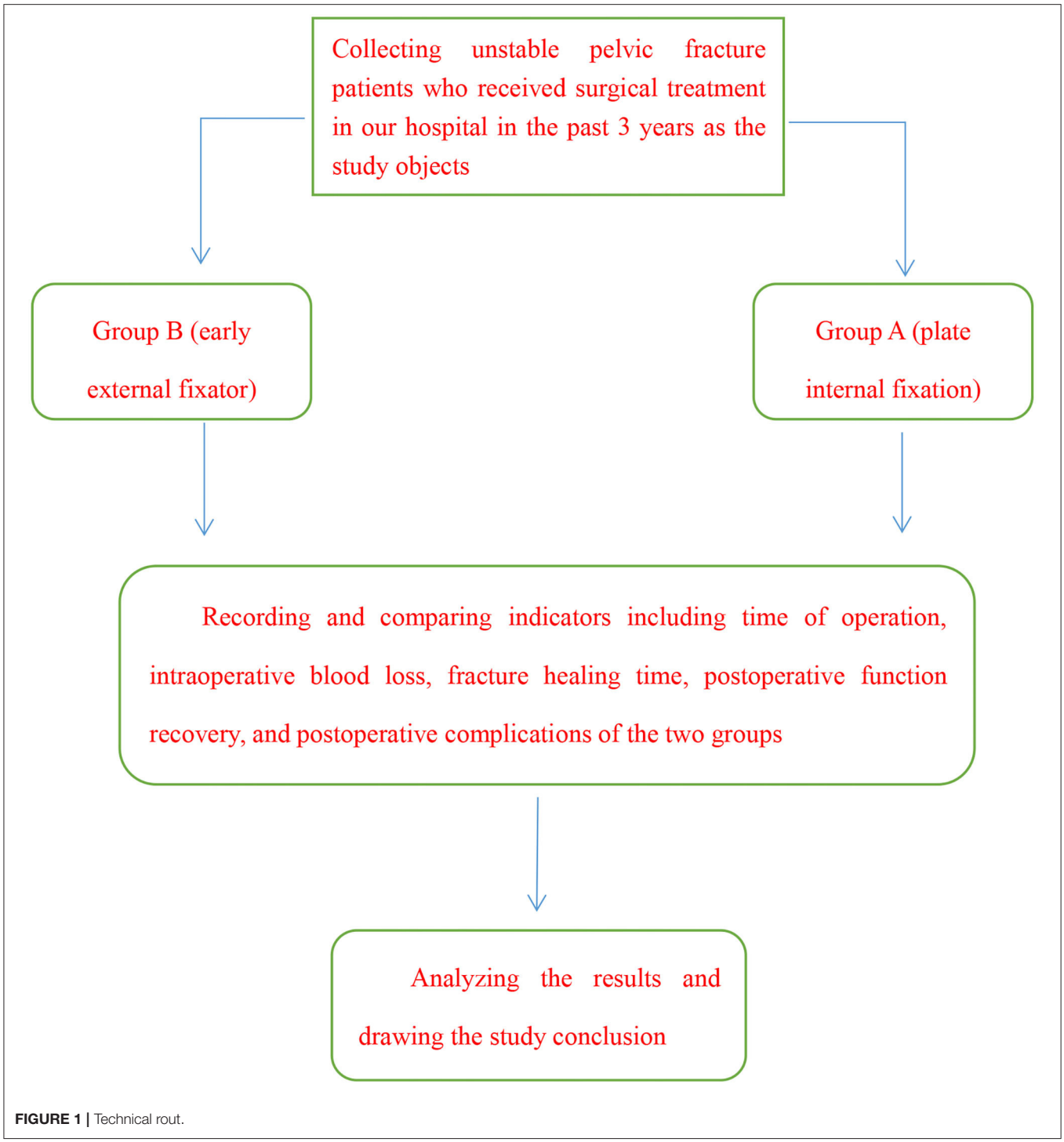


FIGURE 1 | Technical rout.

Patients in group B received early external fixator treatment. After general anesthesia, the patients were in the supine position, routine disinfection and draping were conducted, and before skin incision, 1 g of tranexamic acid (manufacturer: Shandong Yijian Pharmaceutical Co., Ltd.; NMPA approval no. H20043218; specification: 10 ml: 1.0 g) was administered via intravenous infusion. The anterior inferior iliac spine was located by C-arm X-ray, the skin and subcutaneous tissues were cut open, and the lateral femoral cutaneous nerves were protected; intermuscular blunt dissection was performed to expose the anterior inferior iliac spine and place the sleeve and drill holes with a depth over 3 cm (13). After withdrawing the drill, the threaded nails 6 mm in diameter and 150 mm long were screwed in with a depth over 3 cm, and attention should be paid to avoid entering the hip joint.

TABLE 1 | Clinical data ($n = 49$).

Item	Group A	Group B	X^2/t	P
Gender			0.167	0.683
Male/female	27/22	29/20		
BMI (Mean ± SD, kg/m^2)	22.12 ± 1.24	22.43 ± 1.42	−1.169	0.245
Mean age (Mean ± SD, years)	46.61 ± 9.52	45.36 ± 11.54	0.580	0.563
Time injury to surgery (Mean ± SD, d)	6.12 ± 2.26	6.35 ± 2.53	−0.461	0.646
Tile classification			0.400	0.527
B	30 (61.22)	33 (67.35)		
C	19 (38.78)	16 (32.65)		
Cause of injury			0.699	0.873
Traffic accident	23 (46.94)	26 (53.06)		
Fall from height	10 (20.41)	8 (16.33)		
Crushing by weight	12 (24.49)	10 (20.41)		
Others	4 (8.16)	5 (10.20)		
Educational degree			0.238	0.888
College	7 (14.29)	8 (16.33)		
Middle school	28 (57.14)	29 (59.18)		
Primary school	14 (28.57)	12(24.49)		
Place of residence			0.656	0.418
Urban area	25 (51.02)	21 (42.86)		
Rural area	24 (48.98)	28 (57.14)		

When the fracture reduction was satisfied, according to the pelvic 3D-CT image before surgery, the best screw installation position and angle of the external fixator were measured, the connecting rod was installed and the nuts were fixed, and then routine suture was performed.

Observation Indicators

Various perioperative indicators of patients in the two groups were recorded, including time of surgery, intraoperative blood loss, fracture healing time, and total length of the incision.

During postoperative follow-up, reduction evaluation was conducted according to the pelvic imaging data and by Matta's criteria for pelvic ring fractures (14), with displacements of less than 4 mm indicating excellent, 4–10 mm indicating good, 10–20 mm indicating fair, and more than 20 mm indicating poor. The scale could be used to evaluate the quality of fracture reduction, and the good rate = (number of excellent cases + number of good cases + number of fair cases) / total number of cases × 100%.

The patients' hip joint recovery after surgery was assessed by the Harris Hip Scale (HHS) (15), which covered four domains, function, pain, absence of deformity, and range of motion. The maximum score was 100 points, with more than 90 points indicating excellent, 80–90 points indicating good, 70–79 points indicating fair, and less than 70 points indicating poor.

The degree of postoperative pain was assessed by the Visual Analog Scale (VAS) (16), to be specific, a vernier slide caliper about 10 cm long was marked with 10 scales on one side and "0" (no pain) and "10" (the worst pain imaginable) at both ends. In clinical use, the side without the scales was faced to the patients for them to pick the corresponding position that best indicated their current degree of pain, and the doctor gave the score accordingly.

The occurrence of complications after-treatment of the two groups was recorded, including venous thrombosis, neural injury, and incision infection.

Statistical Methods

In this study, the data were processed by the professional statistic software SPSS26.0 (IBM, NY, USA), the picture drawing software was GraphPad Prism 7 (GraphPad Software, San Diego, USA), the enumeration data were examined by X^2 test and expressed by [n(%)], the measurement data were examined by t-test and expressed by Mean±SD, and differences were considered statistically significant at $P < 0.05$.

RESULTS

Clinical Data

No significant between-group differences in clinical data including gender ratio, time from injury to surgery, and cause of injury were observed ($P > 0.05$), presenting comparability. See **Table 1**.

Perioperative Indicators

Various perioperative indicators were significantly better in group B than in group A ($P < 0.001$). See **Table 2**.

Fracture Reduction Quality

The good rate of fracture reduction was significantly higher in group B than in group A ($P < 0.05$). See **Table 3**.

TABLE 2 | Perioperative indicators [Mean ± SD].

Group	n	Time of operation (min)	Intraoperative blood loss (mL)	Fracture healing time (d)	Total length of incision (cm)
A	49	105.47 ± 18.83	102.01 ± 12.26	108.82 ± 8.71	23.05 ± 3.09
B	49	69.81 ± 8.43	75.09 ± 7.35	95.93 ± 4.47	19.11 ± 2.90
t		12.099	13.183	9.214	6.505
P		<0.001	<0.001	<0.001	<0.001

TABLE 3 | Fracture reduction quality [n(%)].

Group	n	Excellent	Good	Fair	Poor	Good rate
A	49	17 (34.69%)	19 (38.78%)	5 (10.20%)	8 (16.33%)	83.67% (41/49)
B	49	23 (46.94%)	17 (34.69%)	7 (14.29%)	2 (4.08%)	95.92% (47/49)
X^2						4.009
P						0.045

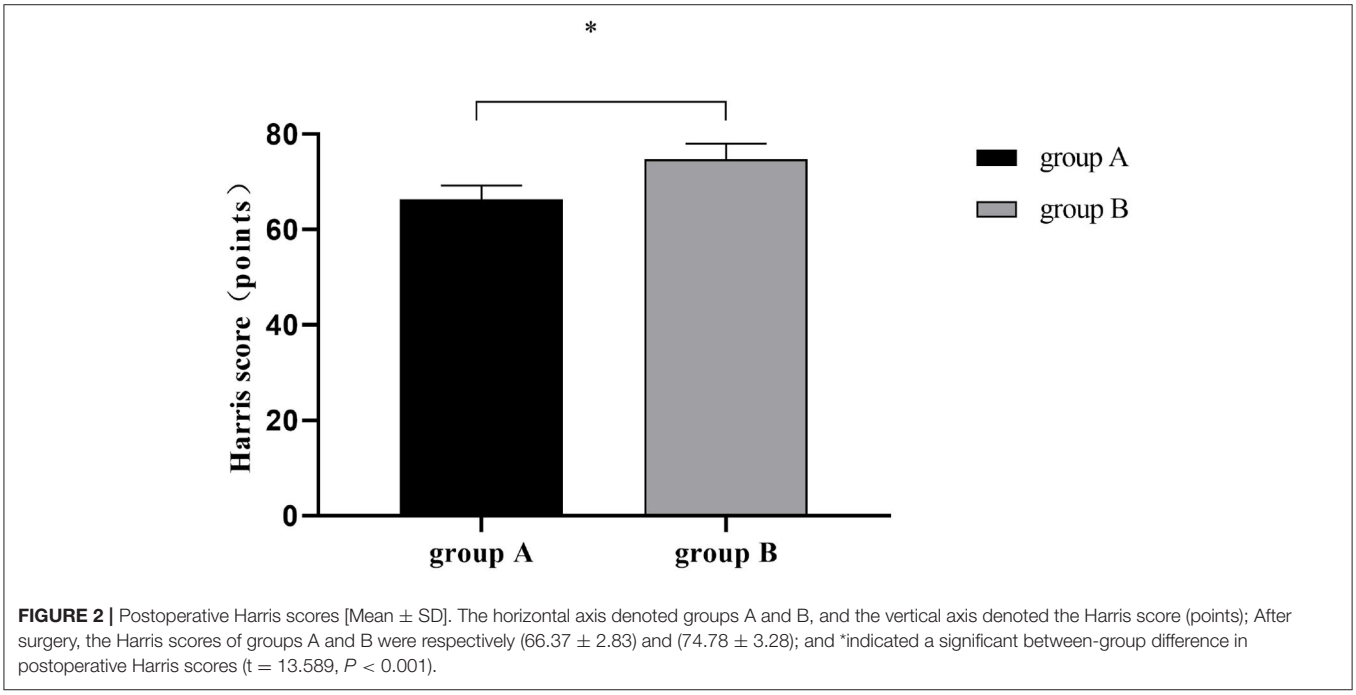


FIGURE 2 | Postoperative Harris scores [Mean ± SD]. The horizontal axis denoted groups A and B, and the vertical axis denoted the Harris score (points); After surgery, the Harris scores of groups A and B were respectively (66.37 ± 2.83) and (74.78 ± 3.28); and *indicated a significant between-group difference in postoperative Harris scores (t = 13.589, $P < 0.001$).

Harris Scores

The postoperative Harris score was remarkably higher in group B than in group A ($P < 0.001$). See **Figure 2**.

VAS Scores

The postoperative VAS score was significantly lower in group B than in group A ($P < 0.001$). See **Figure 3**.

Postoperative Complications

The total incidence rate of postoperative complications was significantly lower in group B than in group A ($P < 0.05$). See **Table 4**.

DISCUSSION

Smart healthcare is a rapidly developing high-tech means of medical treatment in the era of big data, and it has gradually infiltrated into people's daily lifestyles. This technique helps people better solve problems in the field of medical health through artificial intelligence (AI) technology, big data, cloud computing, etc., and then opens up new directions in the construction of medical fields. Each patient is considered as an independent individual in smart healthcare, and multiple diagnosis and treatment modalities depending on different patient conditions can be provided, thereby promoting medical quality. Pelvic fractures are injuries resulting from high energy, which are often accompanied by multiple life-threatening

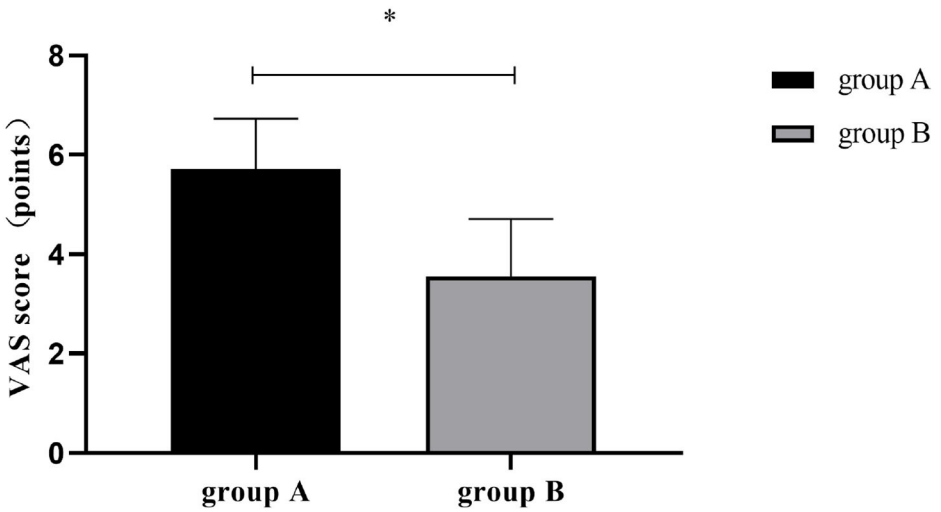


FIGURE 3 | Postoperative VAS scores [Mean±SD]. The horizontal axis denoted groups A and B, and the vertical axis denoted the VAS score (points); After surgery, the VAS scores of groups A and B were respectively (5.71 ± 1.02) and (3.55 ± 1.16); and *indicated a significant between-group difference in postoperative VAS scores ($t = 9.789$, $P < 0.001$).

TABLE 4 | Postoperative complications [n(%)].

Group	n	Venous thrombosis	Nerve injury	Incision infection	Total incidence rate
A	49	2 (4.08)	2 (4.08)	4 (8.16)	16.33% (8/49)
B	49	0 (0.00)	1 (2.38)	1 (2.38)	4.08% (2/49)
X^2					4.009
P					0.045

injuries to other parts of the body, so the implementation of emergency trauma control and restoration of pelvic ring stabilization is essential for patient survival and treatment. Since the deformity caused by pelvic fracture usually triggers long-term dysfunction, surgical reduction and fixation need to be performed to restore the stability of the pelvis (17). Due to the complexity of pelvic structures and fracture types, surgery is often challenging, so there is currently no consensus on the surgical approach and the choice of fixation method. As a simple method of external fixation, traditional complete pelvic fixation has many advantages, such as easy operation and access, making it a common surgical procedure for the treatment of unstable pelvic fractures for a time (18). However, the poor fixation strength and fixation results of this technique make it unsuitable for some complex pelvic fractures. The previously used internal fixation treatment methods have the shortcomings such as a long time of surgery, high difficulty in surgery, long time of postoperative healing, and more postoperative complications. The external fixator technique is frequently applied in the emergency treatment of pelvic fractures and plays a key role in performing reduction, controlling bleeding, and reducing mortality (12). By drawing on previous clinical treatment experience, this study better exerts the application of external fixator technology in a pelvic fracture through a comparative analysis of the clinical effects of early external fixator and plate internal fixation in the treatment of unstable pelvic fracture, presenting a certain reference and guidance significance for the treatment of an unstable pelvic fracture.

The study results showed that the postoperative good rate of fracture reduction was significantly higher in group B than in group A ($P < 0.05$), which was due to the fact that the adjustable external fixation technique can reduce bleeding at the pelvic fracture end and stabilize the condition, providing a surgical opportunity for other combined injuries. In addition, the study also found that, compared with plate internal fixation treatment, an early external fixator can reduce intraoperative blood loss, time of surgery, and total length of incision to a certain extent, and accelerate the fracture healing time, which may be due to the fact that the external fixator treatment causes little skin and muscle damage and only needs to place screws firmly (19), and the efficacy of such easily operative procedure has been demonstrated in multiple metatarsal fractures (20). Moreover, the study results showed that the total incidence rate of postoperative complications was remarkably lower in group B than in group A ($P < 0.05$), and the reason may be that plate internal fixation uses the steel plate to stabilize the anterior and posterior pelvic ring for reconstruction, so as to restore fracture reduction and associated site function with the help of the plate. Some studies have pointed out (21, 22) that plate

internal fixation, although has a good effect, can easily induce damage to soft tissues and nerve tissues and increase the chance of postoperative complications because of the large intraoperative incision and high impact on the body and some soft tissues. Whereas external fixator treatment is highly adjustable because the angulation can be adjusted by the attachment of fixators, which greatly reduces the occurrence of complications and allows the patient to turn over after surgery, alleviates decubital ulcers resulting from prolonged bed rest, and makes nursing more convenient. Meanwhile, it also can increase patient stability during handling and inspection to some extent, and then reduce secondary damage and complications (23, 24). The shortcomings of the study were the small sample size and a single source of cases. So further biomechanical study should be carried out to comprehensively observe the clinical effects of the two different fixation methods, so as to provide an evidence-based basis for the treatment of unstable pelvic fracture patients.

In conclusion, through analysis in the context of intelligent medicine, it is found that early external fixator treatment has better results in treating patients with unstable pelvic fractures. In recent years, with the promotion and improvement of technologies such as minimally invasive surgery and spine-pelvis combined fixation, various new directions have appeared in the treatment of pelvic fracture, meanwhile, with the continuous development of imaging-guided systems and the deepening research of biological models of real pelvic fracture established by computer simulation, conditions have been provided for the broadening of new ideas for clinicians, and the treatment of stable pelvic fractures will continuously develop in the aspects of ease of operation and firmness.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Jiamusi University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HZ and YL: conception and design, data analysis, and interpretation. All authors: administrative support, provision of study materials or patients, collection and assembly of data, manuscript writing, and final approval of manuscript.

FUNDING

This research was funded by Scientific research project of Heilongjiang Provincial Health Commission (Grant No. 2020-327).

REFERENCES

1. Ueda Y, Inui T, Kurata Y, Tsuji H, Saito J, Shitan Y. Prolonged pain in patients with fragility fractures of the pelvis may be due to fracture progression. *Eur J Trauma Emerg Surg.* (2021) 47:507–13. doi: 10.1007/s00068-019-01150-0
2. Chen H, Jiang B, Zhang Q, Chang Z, Yang Z, Zhang J, et al. Intelligent monitoring system assisted pelvic fracture closed reduction for treating one case with complex pelvic fracture. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi.* (2020) 34:874–7. doi: 10.7507/1002-1892.202001035
3. Guillaume JM, Pesenti S, Jouve JL, Launay F. Pelvic fractures in children (pelvic ring and acetabulum). *Orthop Traumatol Surg Res.* (2020) 106:S125–S33. doi: 10.1016/j.otsr.2019.05.017
4. Reinke C, Backer H, Lotzien S, Schildhauer TA, Seybold D, Gessmann J. Arthrodesis of the infected knee joint with the ilizarov external fixator: an analysis of 13 cases. *Z Orthop Unfall.* (2020) 158:58–74. doi: 10.1055/a-0869-8190
5. Yin XL, Tan YR, Liu Y, Sun WW, Zhang XY, Hu YJ, et al. Clinical application of temporary external fixator for immediate mandibular reconstruction. *J Craniofac Surg.* (2019) 30:e337–e42. doi: 10.1097/SCS.0000000000005303
6. Shah FA, Ali MA, Kumar V, Alam W, Hasan O. Does pin tract infection after external fixator limits its advantage as a cost-effective solution for open fractures in low-middle income countries, a prospective cohort study. *J Pak Med Assoc.* (2019) 69:S41–S5.
7. Fuiano M, Mosca M, Caravelli S, Di Liddo M, Grassi A, Majumdar A, et al. Callus distraction with external fixator for the treatment of congenital brachymetatarsia of the fourth ray. *Foot Ankle Surg.* (2020) 26:693–8. doi: 10.1016/j.fas.2019.08.018
8. Mohamad M, Ansorge A, Vieira Cardoso D, Gamulin A, A. case report depicting patient's installation on the fracture table when an ankle spanning external fixator is already in place. *BMC Musculoskelet Disord.* (2019) 20:406. doi: 10.1186/s12891-019-2808-5
9. Dai J, Wang X, Zhang F, Zhu L, Zhen Y. Treatment of distal metaphyseal tibia fractures using an external fixator in children. *Medicine (Baltimore).* (2019) 98:e17068. doi: 10.1097/MD.0000000000017068
10. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* (2013) 310:2191–4. doi: 10.1001/jama.2013.281053
11. Sun ZB, Peng H. Experimental study on the prevention of posttraumatic osteoarthritis in the rabbit knee using a hinged external fixator in combination with exercises. *J Invest Surg.* (2019) 32:552–9. doi: 10.1080/08941939.2018.1543483
12. Chen X, Zhang W, Yu Z, Wang Y, Zhang S, Song T. Application of bone transport with unilateral external fixator combined with locked plate internal fixation in treatment of infected tibial nonunion. *Zhongguo Xiu Fu Chong Jian Wai Ke Za Zhi.* (2019) 33:328–31. doi: 10.7507/1002-1892.201811024
13. Coppa V, Marinelli M, Specchia N. Unilateral uniplanar modular external fixator for percutaneous proximal femoral osteotomy in children: surgical technique. *Eur J Orthop Surg Traumatol.* (2019) 29:205–11. doi: 10.1007/s00590-018-2295-7
14. Roth F, Cagienard F, Link BC, Hodel S, Lehnick D, Babst R, et al. Primary or secondary wound healing of the pin sites after removal of the external fixator: study protocol for a prospective, randomized controlled, monocenter trial. *Trials.* (2020) 21:205. doi: 10.1186/s13063-020-4087-8
15. Kamruzzaman M, Mahboob AH, Saha MK, Islam MS, Alam MT, Alamgir MK, et al. Outcome of ilizarov external fixator for the treatment of gap non-uniting mid shaft tibia-fibula fractures: our experience. *Mymensingh Med J.* (2020) 29:284–9.

16. Thomas M, Hidalgo Diaz JJ, Prunieres G, Facca S, Igeta Y, Liverneaux P. Minimally invasive internal fixation for extra-articular distal radius fracture: Comparison between volar plate and intramedullary nail. *Orthop Traumatol Surg Res.* (2019) 105:409–15. doi: 10.1016/j.otsr.2018.10.013
17. Sheng W, Ji A, Fang R, He G, Chen C. Finite element- and design of experiment-derived optimization of screw configurations and a locking plate for internal fixation system. *Comput Math Methods Med.* (2019) 2019:5636528. doi: 10.1155/2019/5636528
18. Lotzien S, Hoberg C, Rausch V, Rosteius T, Schildhauer TA, Gessmann J. Open reduction and internal fixation of humeral midshaft fractures: anterior versus posterior plate fixation. *BMC Musculoskelet Disord.* (2019) 20:527. doi: 10.1186/s12891-019-2888-2
19. Silva ADM, Furtado G, Dos Santos IP, da Silva CB, Caldas LR, Bernardes KO, et al. Functional capacity of elderly with lower-limb amputation after prosthesis rehabilitation: a longitudinal study. *Disabil Rehabil Assist Technol.* (2021) 16:556–60. doi: 10.1080/17483107.2019.1684581
20. Saruco E, Guillot A, Saimpont A, Di Rienzo F, Durand A, Mercier C, et al. Motor imagery ability of patients with lower-limb amputation: exploring the course of rehabilitation effects. *Eur J Phys Rehabil Med.* (2019) 55:634–45. doi: 10.23736/S1973-9087.17.04776-1
21. Fischer K, Trombik M, Freystatter G, Egli A, Theiler R, Bischoff-Ferrari HA. Timeline of functional recovery after hip fracture in seniors aged 65 and older: a prospective observational analysis. *Osteoporos Int.* (2019) 30:1371–81. doi: 10.1007/s00198-019-04944-5
22. Brand A, Klopfer-Kramer I, Bottger M, Kroger I, Gaul L, Wackerle H, et al. Gait characteristics and functional outcomes during early follow-up are comparable in patients with calcaneal fractures treated by either the sinus tarsi or the extended lateral approach. *Gait Posture.* (2019) 70:190–5. doi: 10.1016/j.gaitpost.2019.03.007
23. Crockett K, Farthing JP, Basran J, Dal Bello-Haas V, Johnston G, Haver CRA, et al. Changes in fall risk and functional status in women aged 50 years and older after distal radius fracture: A prospective 1-year follow-up study. *J Hand Ther.* (2019) 32:17–24. doi: 10.1016/j.jht.2017.09.009
24. Greising SM, Corona BT, Call JA. Musculoskeletal Regeneration, Rehabilitation, and Plasticity Following Traumatic Injury. *Int J Sports Med.* (2020) 41:495–504. doi: 10.1055/a-1128-7128

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Sun and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fpubh.2022.879342

Analyzing the Effect of People Utilizing Mobile Technology to Make Banking Services More Accessible

Jiale Zhu and Manyi Wang*

Capital University of Economics and Business, Beijing, China

Many firms in the modern world utilize m-banking systems to communicate with their consumers. The word m-banking refers to a widespread method of providing financial services and localization to customers. Since m-banking is important to both banks and users, it has been included in numerous literary works. As a result, embracing financial services *via* the m-banking platform is critical. This article's technique is mostly descriptive research that investigates common views, current situations, modern tactics, tangible emerging consequences, etc. The main objective here is to analyze the benefits of this study by investigating the past. Since this article analyzes what exists and is descriptive, the data is being retrieved by conducting a cross-sectional survey method about different features that are relevant by sampling the population. The main aim of this study is to explore the adoption of mobile banking technology by consumers. Based on the values of different variables such as affective commitment (AC), transaction convenience (TC), perceived ease of use (PEU), perceived reliability (PR), pre and post benefits (PPB), service, system, and information quality (SSIQ), bank trust (BT), and profitability (P), the inter-relationship between them and the adoption of m-banking technique by the users in banking technology. The model is investigated by examining the hypothesis and identifying the relationship that exists between these different parameters. A simple linear regression method is implemented using the Statistical Package for the Social Sciences (SPSS) software.

Keywords: mobile technology, banking, financial service, transaction convenience, mobile banking

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Prabhu Kavin,
Sri Ramachandra University, India
Imran Siraj,
St Andrews Institute of Technology & Management, India
Praise Young,
Nnamdi Azikiwe University, Nigeria

****Correspondence:***
Jiale Zhu
zhujiale8899@126.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 February 2022*
Accepted: *17 March 2022*
Published: *29 April 2022*

Citation:
Zhu J and Wang M (2022) Analyzing the Effect of People Utilizing Mobile Technology to Make Banking Services More Accessible. Front. Public Health 10:879342. doi: 10.3389/fpubh.2022.879342

INTRODUCTION

Nowadays, mobile technology plays a vital role in many industries such as banking on managing stocks and finances, business firms sell their products through applications and websites. Mobile banking services such as payment systems influence people's lives more dramatically than other innovations in the recent times (1). Mobile technology made banking services such as payment, transaction, and other services easy through the online mode instead of offline mode. Even the adoption of mobile banking services increased day-by-day by the consumers, still mobile payments do not become ordinary to many of the public (2). Due to the growth of mobile network technology, there is a need for security infrastructure for mobile payments. The security measures give trust to the consumers to adopt mobile banking at anytime from anywhere.

For achieving a successful outcome in business, mobile technology is recognized as competitive and creative equipment that provides better online transaction options (3). Recently, cellphones are being widely used as a tool by people for online shopping, banking, paying bills, etc. Our day-to-day routine lives have been changed through the adoption and development of mobile technology that modifies the way of communication, approaches for selling and buying goods and services, and also the mode of information collection. Such corresponding correlation arises in a constructive environment that has no limitation regarding the time or place (4).

The rapid development in mobile technology and the protection framework facilitate the people to perform m-banking transactions at any time from anywhere, whereas m-banking comprises financial service which is implemented directly or indirectly over a network of wireless telecommunication (3). In the financial sector, the services applied by mobile are referred to as mobile financial services (MFSs) they also include mobile banking and mobile payment. Several studies were conducted to examine the peoples' mindset in view of understanding the mode of adopting mobile banking as a unique and usual service (5). According to the statistics, the level of usage of smartphones has increased from 35 to 80% in 2020, also worldwide, the number of mobile phone beneficiaries is about 6.8 billion, though each mobile was fully activated with the internet. Thus, mobile banking helps in the development of bank access rates which influences bank perception (6). In the banking sector, it is observed that most of the customers have switched to smartphone, apps, or tablets which are highly utilized in their day-to-day life for shopping, entertainment, learning, socializing, and jobs, and these mobile-centric customers are preferred by more bankers (7). For example, a bank in Japan, Jibun Bank whose principal communication with customers is through the mobile channel and was the first bank that affords their complete banking services and products over mobile channels. With the help of cameras and mobile phone, customer has the facility to start a new account with their bank. This mode of banking enables the bank to improve their customer service over anywhere at any time basis. Also, the different mobile banking services are payment, web-related shopping, fund transfer, and web-related marketing. This Jibun Bank has reached more popular for its efficient services within the short duration of their mobile banking introduction (8). Through this, they have attained above 500,000 new bank accounts in their bank. In the initial days, the bank had the facility to send usual text messages related to money transfer and withdrawal and these services were updated to provide its customer a complete functional banking experience (9).

Mobile banking is adopted by many consumers as it is easy to open an account, pay bills, transfer funds, do mobile-based shopping, etc. Many banks in China and America who introduced mobile banking for all their services experienced high demand for their services through mobile banking (10). But many challenges were faced by these banks due to no face-to-face interactions. Trust is an important factor in mobile banking but lack of interaction led to the risk of uncertainty and anonymity. On the other hand, the consumers also suffer from uncertainty whether the bank is trustworthy are not.

During the last decades, mobile banking became relevant all over the world. The number of fund transfers through mobile banking also increased since last year, but mobile banking has many security problems. Sometimes, the security setting of mobile phones can be overridden by the virus in smartphones (11). Many banks open their banking applications, these apps should be updated frequently otherwise the consumers can be vulnerable to the attacks such as DDoS attacks, phishing, spoofing, corporate account takeover, and skimming (12).

The major contribution of the article is:

- To analyze the benefits of this study by investigating the past.
- To explore the adoption of mobile banking technology by consumers.

The remaining article is arranged accordingly. Section Literature Review presents the related work and the hypothesis is formulated in Section Hypothesis Development and Research Methodology. The results in terms of linear regression and descriptive analysis are provided in Section Results and the article is concluded in Section Conclusion and Discussions.

LITERATURE REVIEW

In the modern world, many firms use systems of mobile banking to communicate with the customers. The payment system in mobile banking services benefited people's lives. This article aims to analyze the effect of people using the technology of mobile to make banking services more accessible. Literature on mobile banking adoption, electronic banking services, mobile payment service user acceptance are reviewed.

Hamidi et al. (13) studied the influence of mobile banking adoption on consumer engagement and satisfaction utilizing the customer relationship management (CRM) system, which is the most essential aspect in banking industry. CRM is also seen as a critical role for enhancing client satisfaction in mobile banking. The statistical study performed evaluated the dialogue between the bank's customer sector and their client. The statistical analysis findings have a favorable influence on consumer interactions and satisfaction.

Geebren et al. (14) investigated the importance of consumer satisfaction in mobile eco-systems that used electronic banking services, particularly in developing nations. This entailed researching consumer satisfaction in mobile banking, with a focus on the importance of trust. To determine consumer satisfaction, structural modeling using partial least squares (PLS–SEM) methods were employed to examine the data, and trust demonstrated that customer contentment had a beneficial influence.

The relevance of an early trust theoretical model in mobile payment service user acceptance was highlighted by Gao et al. (15). The initial trust theoretical model highlighted the facilitators and barriers to user trust in m-payment services. The links in the original trust theoretical model were assessed using partial least squares structural modeling (PLS–SEM). The

findings may be used in m-payment adoption research and practice in a variety of ways. In total, 52.3% of the difference in usage intention was explained by the current model.

Hentzen JK et al. (16) offered a mobile technology that allows for vital involvement, as well as an explanation of how a retirement app might assist people in planning for their post-retirement strategy. The available literature survey data were used to evaluate a sample of 440 Australian pension fund members. The findings show that consumers' perceived financial security, financial self-efficacy, retirement planning involvement, future consequences consideration, and perceived usefulness with a mobile retirement app have direct and indirect effects on their expected engagement through their goal to adopt the app.

Zhu et al. (17) investigated existing technology designs, including mobile banking, used by rural communities in six Chinese regions. According to the findings, interpersonal and mass communication channels have a bigger influence than organizational communication channels. Mobile banking should be examined since it can assist alleviate the lack of access to financial goods and financial infrastructure in rural areas.

Afeti et al. (18) developed a mobile payment technology for payment for micro-businesses. The study draws on the transaction cost theory and the task-technology fit (TTF) theory as the assumed lens. In total, 20 micro-businesses based on qualitative data were analyzed and the research findings denote those micro-businesses adoption of mobile payments results in strategic and operational benefits.

Alamoudi et al. (19) proposed that the mobile technology acceptance lookalike was changed as we investigated consumers' acceptance of mobile shopping in general stores by examining transaction convenience, usefulness, attitudes, ease of use, transaction speed, optimism, and personal innovativeness. A total of 351 respondents completed the questionnaire evaluation. Consumers are willing to use mobile shopping channels if the system is clear and straightforward to use.

Jebarajakirthy et al. (20) used a comprehensive moderated mediation framework to evaluate the influence of online convenience aspects on mobile banking uptake. Covariance-based structural equation modeling and the process macro are utilized to test these predictions. This study examines how convenience characteristics influence mobile banking adoption intentions.

Abdinoor et al. (21) studied the adoption of mobile financial services in Tanzania with the use of a technology acceptance model. To select the sample from data collection, a random sampling technique was used. The user and non-user of mobile financial services were included in the sample. Zhang et al. (3) investigated customers' use of mobile technology to help them with banking services and activities, as well as the variables that impact their adoption and engagement. Here, the analysis is done by the structural equation modeling technique to know the consumers' intentions toward mobile banking. The result examines the adoption of mobile banking apps to facilitate bank consumers' banking services.

Tiwart et al. (22) researched the variable that influences adoption in Commercial Bank Ethiopia. The author used factors such as perceived ease of use (PEOU), infrastructure (INF), security (SEC), trust, and e-banking adoption. Here, a structural equation model based on least partial square analysis is used. The result of this analysis proves that the trust of customers mediates between PEOU, INF, SEC, and e-banking adoption and realized the factors that guess the purpose and adoption of e-banking by Ethiopian customers.

The problem identified in mobile banking services is critical financial services through mobile banking because it has security problems.

HYPOTHESIS DEVELOPMENT AND RESEARCH METHODOLOGY

Participant's Demographic Information

This study is mainly conducted in China by taking the top e-banks that utilize m-banking. The questionnaire is distributed to both the users and staff of the bank. Within 6 months, 293 questionnaires were completed. These details were taken into consideration in this study. The details of the questionnaires were obtained online and also provided to the users on their official bank pages. The information obtained in the questionnaire is provided in **Tables 1**, **2**.

TABLE 1 | Demographic analysis.

Demographics variables	Percentage	Number
Sex		
Female	54.50	121
Male	45.49	101
Age		
1–20	3	25
20–30	59	120
30–40	15	14
>40	23	63
Education		
Master of science	12	25
Bachelor of science	15	32
Engineering	20	57
Medicine	18	60
Diploma	40	85
Associate degree	17	14
PhD	5	7

TABLE 2 | Participant details.

Bank types	User rating
State-owned	3.25
Joint-stock	2.14
Postal savings	3.25
Agricultural bank	2.52
Commercial city bank	1.54

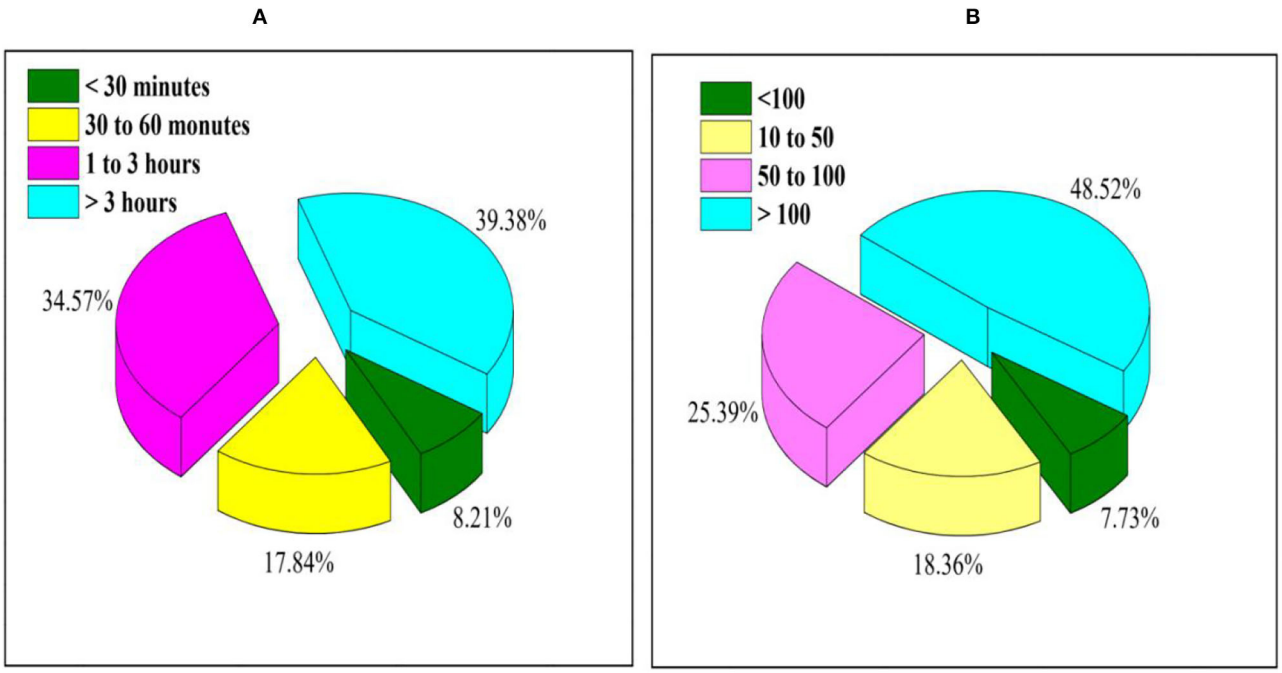


FIGURE 1 | Usage analysis of m-banking application **(A)** Users using m-banking during the day and **(B)** users' friends and references using mobile banking.

Participant Details

The information of the customers is mainly obtained based on their m-bank usage. The five main banks such as state-owned, joint-stock, postal savings, commercial city bank, and the agricultural bank were selected. Each participant was asked to say about their experience of their usage.

The amount of time spent by each user on average is provided in **Figure 1A**. As seen in **Figure 1A**, more than 39.38% of users are using more than 3 h a day and a minimum of is online Conceptual Model for less than 1 h. In total, 8.21% of the users only use the application for a limited amount of time (below 30 min).

Figure 1B presents the percentage of people the users have referred for mobile banking. Based on **Figure 1B**, a total of 48.52% has more than 100 contacts who have been referred to the m-banking. The increase in m-banking contact can improve efficiency.

Conceptual Model

The main aim of this research is to identify the usage of mobile technology of consumers for the banking system. The hypothesis provided for each variable is explained in detail in this section and the conceptual model is provided in **Figure 2**. The measurement instrument used for this study is presented in **Table 3**.

Hypothesis Modeling

Service, System, and Information Quality (SSIQ)"

System quality mainly measures the efficiency of the overall m-banking system in terms of service provider anonymity. The system quality of the m-banking mainly depends upon the ease of use, the attractiveness of the application, trust of the user, etc. If the user trusts the system quality, then they mainly select the system for future transactions. Service Quality mainly relies on the service provided by the m-banking system and it also evaluated the effectiveness of the service in terms of personalization, reliability, delay, etc. Hence, service quality is crucial to improve the quality of m-banking since poor-service quality can result in minimal user trust and satisfaction. Information quality mainly implies that high-quality information is reliable, complete, accurate, relevant, and accessible. The information quality is an important variable that improves the usage, trust, and satisfaction of the customers in the m-banking system.

Hypothesis (H6): The service, system, and information quality had a positive relationship in improving the m-banking quality and user satisfaction.

Affective Commitment

Affective obligations between business partners reflect mental reliance on others and are founded on emotions, loyalty, and dependency. The psychological bond in this example demonstrates an emotional commitment. Support for differences in particular data among new and long-term clients. Based on their differing expectations, new and loyal clients behave in the same way.

Hypothesis (H1): The interaction of customers helps to develop a good relationship with the affective commitment of a person to the bank which is focused on interaction.

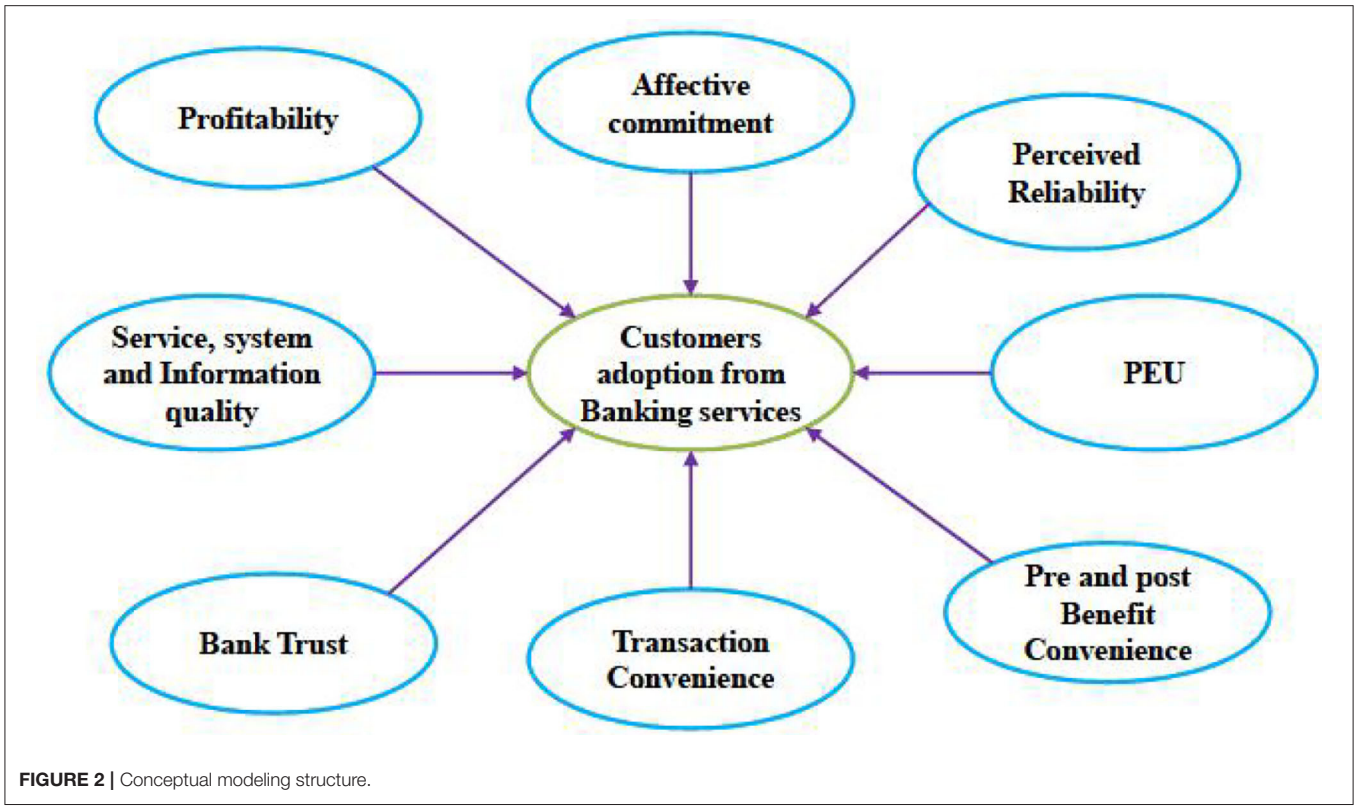


FIGURE 2 | Conceptual modeling structure.

Pre- and Post-benefit Convenience

The time and attempt to acquire some benefits of services is said to be benefit convenience. Some of the components of benefit convenience are fast service, timely services, and bank employee attitudes. Sometimes in banks, the consumer needs to visit repeatedly to avail specific service. But in m-banking, consumers can avail any services on the go from the home itself. So, the time and attempt to acquire the service became very easy. Therefore, benefit convenience shows a positive result in mobile banking adoption (23). The time and attempt to contact a service provider for a particular service is said to be post benefit convenience. In m-banking, the service provider can be contacted and solve the grievances within a few clicks through various means such as email, live chat, and toll-free numbers. But in offline banking, the consumer needs to visit the branch and wait for the service provider to solve the problem in service. So, the post benefit convenience shows a positive result in mobile banking adoption (20).

Hypothesis (H5): The pre- and post-benefit convenience are positively interrelated with m-banking.

Transaction Convenience

The fast and simple way to complete the transaction is said to be transaction convenience. Acquiring the service of transaction in a lesser time is the main component of transaction convenience. Other components of transaction convenience are uninterrupted transaction, transaction confirmation, easy checkout, and price inconsistency. By using m-banking, the transaction can be done in any place at any time in a few clicks and can do several transactions simultaneously. But in offline banking, consumer needs to wait in a queue for availing a transaction service. Hence, transaction convenience shows a positive result in mobile banking adoption (24).

Hypothesis (H2): The transaction convenience is positively interrelated with m-banking.

Profitability

Customer participation in an organization's mobile banking may increase customer value and profitability (25).

Hypothesis (H8): Profitability shows positive results in mobile banking adoption.

Perceived Reliability

Perceived reliability is specified as the limit to which the individual has independently considered the technology as a faithful one. In order to offer a trustworthy technical service, the system reliability is important that enables the user to attain the aimed objective. An undependable technology that is employed by an individual seems to be low confidence and such technology utilization should be restricted (26). Research related to mobile technology and information system has revealed that a technology with perceived reliability possesses has an important impact on customers' perceived satisfaction, detected value, and observed quality. Also, customers' understanding level of trust is examined and analyzed as an antecedent of confidence in an organization through the mobile commerce setting. Though mobile is a wireless technology, it is easily susceptible to

TABLE 3 | Measurement instrument to analyze different parameters.

Variables	Measurement instrument
Bank trust	The bank helps me to use the mobile device of the consumer for accessing the services BT1: The m-banking is trustworthy BT2: The m-banking mainly satisfies the promises and commitments
Pre and post benefit convenience	The m-banking resolves the problems in a fast manner Obtaining the follow-up service is easy The solutions provided by m-banking is fast and reliable The banking service is provided with little effort The services are easier to avail The problem is resolved fastly The time taken by it is reasonable
Service, system, and information quality (SSIQ)	Convenient access of m-banking services Easy navigation of m-banking services Visual attractiveness of m-banking services Accurate information of m-banking services Up-to-date information of m-banking services Dependable services provided Personalized m-banking services Professional m-banking services Timely m-banking services
Transaction convenience	Easier to complete the transaction Fast service access Little effort to complete the transactions
Perceived ease of use	Usability of mobile devices for banking Interaction with the mobile devices for banking Skillset achieved in accessing mobile devices for banking
Perceived reliability	To select the supporting technology for mobile banking Ensuring that the technology supporting m-banking does not fail The technology used is reliable
Portability	Increased customer interaction Increased customer visit

violations or attacks (27). Mostly, the customers were insisted to share their personal data while consuming and shopping services performed through mobile technology such as date of birth, debit and credit card details, funds, and address. The reliability of mobile technology should be improved thereby increasing customers' belief in securing their personal data. The postulated hypotheses are mentioned as follows:

Hypothesis (H4): Perceived reliability in m-banking helps to improve the consumer's trust and ease of use of banking services.

Banking Trust

In this hypothesis, trust is considered as a faith of competence, integrity, ability, and benevolence that the individual has toward each other. Trust has become an important reducing feature of risk, whereas the e-commerce connection is considered as a naturally risky factor (28). Also, trust in banking is specified that it is linked to reliability and perceived privacy. Generally, the primary interaction between the customer and banker will significantly affect the trust progress in them.

Hypothesis (H7): Based on their usage of mobile devices and attitude, the consumer trust toward M-banking increases.

Perceived Ease of Use (PEU)

Perceived ease of use evaluates the increase in the amount of work required to acquire a job utilizing new technologies on an individual basis. In the effort condition, the simplicity of use of mobile innovation is critical for including customers in the co-creation of value experiences (29). Mobile technology's ability to increase advantages for bank clients has been widely shown, and its simplicity of use is a key factor in consumer acceptability. PEU has been linked to advanced tackling new technologies in the banking industry in several research. However, the launch was difficult, and it was claimed that customers' inclination to utilize mobile banking services is not intrinsically related to employment prospects. As a consequence, it indicates that study into PEU and its influence on consumer attitudes, as well as the interactive method, are both worthwhile endeavors (30–32).

Hypothesis (H3): Customers' views about the usage of mobile devices for facilitating financial services are favorably associated with their perceived ease of use (PEU).

RESULTS

Scale Validity and Reliability

The reliability of the scales was assessed using Cronbach's alpha values. Goodness-of-fit approaches were used to assess the overall model fit for confirmatory factor analysis (CFA). The RMSEA was 0.09, the comparative fit index (CFI) was 0.99, the goodness-of-fit (GFI) was 0.98, and the standardized root mean square residual (RMR) was 0.06. The X2 to df ratio was 4.14, the root mean square error of approximation (RMSEA) was 0.09, the CFI was 0.99, the GFI was 0.98, and the standardized RMR was 0.06. As indicated in **Table 4**, the reliability coefficients of all constructs were more than 0.87, which is higher than the 0.9 thresholds. The extracted average variance was used to assess convergent validity (AVE). Because the AVE values ranged from 0.58 to 0.92, convergent validity was not an issue. The AVE scores were lower than the squared correlations between pairs of constructs, indicating discriminant validity in the data set. **Table 4** shows the results of the Cronbach's alpha and AVE tests.

Linear Regression Test

Regression tests are run when the allowable alpha coefficient for the scales has been determined. **Table 5** shows the regression test results for each of the assumptions discussed in the following sections. Affective commitment (AC) and interaction were investigated, and the correlation coefficient 0.297 and likelihood of significance 0.000 indicate that there is a positive and significant relationship between the two. The regression test supports this positive relationship, showing that the interaction variable may be able to predict AC variation.

With a correlation value of 0.333 and a probability of significance of 0.000, the following test demonstrated a positive relationship between perceived ease of use and satisfaction (PEU). The correlation coefficient between the variables perceived reliability (PR) and interaction was 0.172, with a

TABLE 4 | Hypothesis testing.

Variables	Cronbach's alpha	Average variance extracted (AVE)	AC	TC	PEU	PR	PPB	SSIQ	BT	P
Affective commitment (AC)	0.78	0.58	0.77							
Transaction convenience (TC)	0.96	0.87	0.55	0.94						
Perceived ease of use (PEU)	0.97	0.88	0.62	0.68	0.95					
Perceived reliability (PR)	0.98	0.91	0.51	0.52	0.56	0.96				
Pre and post benefits (PPB)	0.98	0.92	0.55	0.74	0.66	0.53	0.97			
Service, system, and information quality (SSIQ)	0.95	0.82	0.47	0.56	0.66	0.44	0.74	0.92		
Bank trust (BT)	0.87	0.68	0.72	0.46	0.48	0.52	0.49	0.42	0.83	
Profitability (P)	0.83	0.64	0.67	0.42	0.44	0.50	0.46	0.39	0.80	0.84

TABLE 5 | The results of the linear regression test.

The dependent variable	Independent variable	Number theory	Test significance *P*-value	The correlation coefficient R^2	Result
Affective commitment (AC)	Interaction	H1	0.000	0.297	Accepted
Transaction Convenience (TC)	Interaction	H2	0.000	0.339	Accepted
Perceived ease of use (PEU)	Satisfaction	H3	0.000	0.333	Accepted
Perceived reliability (PR)	Interaction	H4	0.000	0.172	Accepted
Pre and post benefits (PPB)	Interaction	H5	0.000	0.242	Accepted
Service, system, and information quality (SSIQ)	Satisfaction	H6	0.000	0.430	Accepted
Bank trust (BT)	Interaction	H7	0.000	0.347	Accepted
Profitability (P)	Interaction	H8	0.000	0.230	Accepted

probability of significance of 0.000, demonstrating a positive link between the two variables. The correlation value of 0.430 and the probability of significance of 0.000 indicated a positive relationship between satisfaction and service, system, and information quality (SSIQ) (3).

Descriptive Data

The items in this questionnaire were answered using a five-point Likert scale, with the alternatives being: very low (1), low (2), medium (3), high (4), and very high (5). **Table 6** shows the SD, middle value (MV), and lowest value (LV) of the descriptive data received from the surveys (3).

CONCLUSION AND DISCUSSIONS

Mobile banking is a cost-effective way to reach customers. The different activities that can be performed by the consumers in m-banking are checking their bank balance, making transactions, making investments, getting account statements, paying bills, etc. The motivation for user adoption and use of mobile banking is examined in this research. Positive connections between profitability, trust, and transaction convenience to utilize mobile banking have been found in previous studies. Users' confidence in using mobile banking services, enjoyment, utility of the system, etc., are not the same thing. The participants were unconcerned

TABLE 6 | Descriptive data analysis.

Variables	Standard deviation (SD)	Middle	Highest	Lowest
Affective commitment (AC)	1.208	2.52	5	1
Transaction convenience (TC)	1.081	3.36	5	1
Perceived ease of use (PEU)	1.061	3.21	5	1
Perceived reliability (PR)	1.689	2.92	5	1
Pre and post benefits (PPB)	1.393	2.67	5	1
Service, system, and information quality (SSIQ)	1.206	2.79	5	1
Bank trust (BT)	0.955	3.31	5	1
Profitability (P)	1.220	3.10	5	1

about the danger of fraud, system dependability, or perceived privacy while building and extending faith in their banks and mobile services. Because of today's technology-driven lifestyle, individuals are receptive to embracing new technologies that are compatible with their mobile phones.

To add value to practice, we must obtain convincing and widespread results from both users and mobile banking service providers. Like mobile phones, smartwatches have become more popular among people nowadays. With this wearable technology, banks have the greatest responsibility to fulfill hedonic and social needs. The purpose and use of this research are common to all consumers irrespective of various countries, populations, and socio-economic status. The variables used in this work are, namely, affective commitment (AC), transaction convenience (TC), perceived ease of use (PEU), perceived reliability (PR), pre- and post benefits (PPBs) service, system, and information quality (SSIQ), bank trust (BT), and profitability (P). Consumer behavior such as untrustworthiness, fraudulence, and so on, as well as mobile banking transaction risks such as financial, privacy, and cyber security, are important risk concerns that may be addressed in future research developments.

The problems and risks that arise by the consumers and banking sector for adopting mobile banking services would be reduced by the scientific researchers using various technologies and methodologies in the future. The results in this article show the different statistical surveys and believed that it is true for both mobile banking consumers and non-mobile banking consumers with various socioeconomic environments. But when we consider globally each country had its own rules and regulations regarding mobile banking transactions, services, and products. And these rules and regulations react differently in different countries.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Patil P, Tamilmani K, Rana NP, Raghavan V. Understanding consumer adoption of mobile payment in India: Extending Meta-UTAUT model with personal innovativeness, anxiety, trust, and grievance redressal. *Int J Inf Manage.* (2020) 54:102144. doi: 10.1016/j.ijinfomgt.2020.102144
2. Martin J. 7 reasons mobile payments still aren't mainstream. (2016).
3. Zhang T, Lu C, Kizildag M. Banking "on-the-go": examining consumers' adoption of mobile banking services. *Int J Quality Serv Sci.* (2018) 10:279–95. doi: 10.1108/IJQSS-07-2017-0067
4. Sharma SK, Al-Muharrami S. Mobile banking adoption: key challenges and opportunities and implications for a developing country. In: *Emerging Markets from a Multidisciplinary Perspective*. Springer, Cham. (2018). p. 75–86. doi: 10.1007/978-3-319-75013-2_7
5. Riquelme HE, Rios RE. The moderating effect of gender in the adoption of mobile banking. *Int J Bank Market.* (2010) 28:328–41. doi: 10.1108/02652321011064872
6. Sanakulov N, Karjaluoto H. Consumer adoption of mobile technologies: a literature review. *Int J Mobile Commun.* (2015) 13:244–75. doi: 10.1504/IJMC.2015.069120
7. Shareef MA, Baabdullah A, Dutta S, Kumar V, Dwivedi YK. Consumer adoption of mobile banking services: An empirical examination of factors according to adoption stages. *J Retail Consumer Serv.* (2018) 43:54–67. doi: 10.1016/j.jretconser.2018.03.003
8. Hanafizadeh P, Behboudi M, Koshksaray AA, Tabar MJS. Mobile-banking adoption by Iranian bank clients. *Telemat Inform.* (2014) 31:62–78. doi: 10.1016/j.tele.2012.11.001
9. Srivastava M, Fernandes S. An empirical study to measure consumer adoption of mobile banking services. *Int J Public Sector Perform Manage.* (2022) 9:165–89. doi: 10.1504/IJPSPM.2022.119834
10. Narteh B, Yeboah-Asiamah E, Mackin EA. Analysis of young banked and unbanked customers' usage, satisfaction, trust and loyalty for mobile money services in Ghana. *Int J Bus Syst Res.* (2022) 16:40–64. doi: 10.1504/IJBSR.2022.119602
11. Nur T, Panggabean RR. Factors influencing the adoption of mobile payment method among generation Z: the extended UTAUT approach. *J Account Res Organiz Econ.* (2021) 4:14–28. doi: 10.24815/jaroe.v4i1.19644
12. Muhal MA, Luo X, Mahmood Z, Ullah A. Physical unclonable function based authentication scheme for smart devices in Internet of Things. In: *2018 IEEE International Conference on Smart Internet of Things (SmartIoT).* (2018) p. 160–165. IEEE. doi: 10.1109/SmartIoT.2018.00037
13. Hamidi H, Safareeyeh M. A model to analyze the effect of mobile banking adoption on customer interaction and satisfaction: a case study of m-banking in Iran. *Telemat Inform.* (2019) 38:166–81. doi: 10.1016/j.tele.2018.09.008
14. Geebren A, Jabbar A, Luo M. Examining the role of consumer satisfaction within mobile eco-systems: Evidence from mobile banking services. *Comput Human Behav.* (2021) 114:106584. doi: 10.1016/j.chb.2020.106584

15. Gao L, Waechter KA. Examining the role of initial trust in user adoption of mobile payment services: an empirical investigation. *Inf Syst Front.* (2017) 19:525–48. doi: 10.1007/s10796-015-9611-0
16. Hentzen JK, Hoffmann AO, Dolan RM. Which consumers are more likely to adopt a retirement app and how does it explain mobile technology-enabled retirement engagement? *Int J Consumer Stud.* (2021) 46:368–90. doi: 10.1111/ijcs.12685
17. Zhu Q, Lyu Z, Long Y, Wachenheim CJ. Adoption of mobile banking in rural China: Impact of information dissemination channel. *Socio-Econ Plann Sci.* (2021) 101011. doi: 10.1016/j.seps.2021.101011
18. Afeti EY, Owusu A. Impact of mobile payments on micro-business activities: a developing country experience. In: *Digital Innovations, Business and Society in Africa*. Springer, Cham. (2022). p. 75–95. doi: 10.1007/978-3-030-77987-0_4
19. Alamoudi H. Examining Retailing Sustainability in the QR Code-Enabled Mobile Payments Context During the COVID-19 Pandemic. *Int J Customer Relationship Market Manage.* (2022) 13:1–22. doi: 10.4018/IJCRMM.289210
20. Jebarajakirthy C, Shankar A. Impact of online convenience on mobile banking adoption intention: a moderated mediation approach. *J Retail Consumer Serv.* (2021) 58:102323. doi: 10.1016/j.jretconser.2020.102323
21. Abdinoor A, Mbamba UO. Factors influencing consumers' adoption of mobile financial services in Tanzania. *Cogent Busin Manage.* (2017) 4:1392273. doi: 10.1080/23311975.2017.1392273
22. Tiwari P. Electronic banking adoption in Ethiopia: an empirical investigation. *SN Bus Econ.* (2021) 1:1–28. doi: 10.1007/s43546-021-00114-0
23. Shen YC, Huang CY, Chu CH, Hsu CT. A benefit–cost perspective of the consumer adoption of the mobile banking system. *Behav Inf Technol.* (2010) 29:497–511. doi: 10.1080/01449290903490658
24. Trialih R, Astuti ES, Azizah DF, Mursityo YT, Saputro MD, Aprilian YA, et al. How mobile banking service quality influence customer satisfaction of generation x and y? In: *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. (2018) p. 827–32. IEEE. doi: 10.1109/ICTC.2018.8539720
25. Gupta SD, Raychaudhuri A, Haldar SK. Information technology and profitability: evidence from Indian banking sector. *Int J Emerg Market.* (2018) 13:1070–87. doi: 10.1108/IJoEM-06-2017-0211
26. Prabhu Kavin B, Ganapathy S. EC(DH)2: an effective secured data storage mechanism for cloud based IoT applications using elliptic curve and Diffie-Hellman. *Int J Internet Technol Secur Transact.* (2020) 10:601–17. doi: 10.1504/IJITST.2020.10029366
27. Prabhu Kavin B, Ganapathy S. A new digital signature algorithm for ensuring the data integrity in cloud using elliptic curves. *Int Arab J Inf Technol.* (2021) 18:180–90. doi: 10.34028/iajit/18/2/6
28. Prabakaran S, Ramar R, Hussain I, Kavin BP, Alshamrani SS, AlGhamdi AS, et al. Predicting attack pattern via machine learning by exploiting stateful firewall as virtual network function in an SDN network. *Sensors.* (2022) 22:709. doi: 10.3390/s22030709
29. Ramachandran V, Ramalakshmi R, Kavin BP, Hussain I, Almaliki AH, Almaliki AA, et al. Exploiting IoT and its enabled technologies for irrigation needs in agriculture. *Water.* (2022) 14:719. doi: 10.3390/w14050719
30. Adam MNM, Adam MNM, Idris AAA, Ali ASM, Khalid IO, Evaluation of the effect of time of fixation and microwave treatment on quality of fatty tissue fixation in breast cancer specimens. *SPR.* (2021) 2:445–51. doi: 10.52152/spr/2021.166
31. Mshvidobadze T. Bioinformatics as emerging tool and pipeline frameworks. *SPR.* (2021) 2:361–5. doi: 10.52152/spr2021.156
32. Dani R, Juyal D, Rawal YS. A Critical analysis of the restaurant industry's effect on environment sustainability. *SPR.* (2021) 2:385–92. doi: 10.52152/spr/2021.159

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 02 May 2022
doi: 10.3389/fpubh.2022.879530

The Mediating Role of New Media Engagement in This Digital Age

Zhang Yujie[1], *Megat Al Imran Yasin*[1]*, *Syed Agil Bin Shekh Alsagoff*[1] and *Ang Lay Hoon*[2]

[1] Department of Communication, FBMK University Putra Malaysia, Serdang, Malaysia, [2] Department of Foreign Languages, FBMK University Putra Malaysia, Serdang, Malaysia

New Media Technologies refer to all kinds of web-related technologies like social networking sites, blogs, online social media networking, and other communication technology forms. The primary purpose of this study is to examine the mediating role of new media engagement in this digital age. Furthermore, this study defines the concept of new media, highlights the characteristics of new media technologies, and analyzes the application of new media technologies in this digital age. To meet the objectives stated, the researchers conduct a questionnaire survey to showcase that new media engagement plays a mediating role between the instructions provided to the newsroom in this digital age and the attitude of people toward social media. Results of the study state that the majority of the individuals in this digital era display positive engagement with new media technologies. People in this age believe that social media is safe and less risky for all individuals with regard to communication and interactions.

Keywords: new media technologies, social networking, digital age, communication technologies, attitude toward social media

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Prabhu Kavin,
Sri Ramachandra University, India
Praise Young,
Nnamdi Azikiwe University, Nigeria
Imran Siraj,
St. Andrews Institute of Technology & Management, India

***Correspondence:**
Megat Al Imran Yasin
megatalimranyasin@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 February 2022*
Accepted: *22 March 2022*
Published: *02 May 2022*

Citation:
Yujie Z, Al Imran Yasin M, Alsagoff SABS and Hoon AL (2022) The Mediating Role of New Media Engagement in This Digital Age. Front. Public Health 10:879530. doi: 10.3389/fpubh.2022.879530

INTRODUCTION

In recent years, there has been growing interest in investigating the role new media technologies play in diverse areas including business, media, education, mass communication, and so on (1, 2). New media technologies are popularly referred to as Web 2.0 technologies. These technologies use web technologies like online social networking, blogs, wikis, and other social media platforms to communicate with a larger group audience. The last few years have witnessed a steep decline in the use of old media such as magazines, television, newspapers, radio, and other traditional media forms, while the internet has taken over the place of old media rapidly; and for that very reason, old forms of mass communication are becoming extinct with each passing day.

Several studies, for instance Lysak et al. (3), Andrus (4), and Boulos et al. (5) have demonstrated that television took the place of life magazines; emails and telephone took over the telegraph; and so forth. From these instances, it can be understood that old media has been replaced by modern communication techniques. New media communication technologies supplemented the different forms of old media by working in parallel to meet the demands of different organizations. Nowadays, any kind of mass communication is available on the internet and at newsstands (5). Mass communication media allows the readers to comment in real-time on blog sites and online social networking sites. Developments in the arena of new media technology permit individuals to participate in public discussions (6). The whole of mass communication technology has changed with the recent developments in the field of technology.

An entirely new shape has been given to the communication process where people began to relate to new media on a personal level unlike the old media forms (7). Many researchers stressed the fact that the concept of virtual identity is growing popular in recent years, as it will help them to understand the role of new media technologies in developing one's own identity in this digital age (7, 8). This will also help the researchers to analyze how people shape their own identity by communicating through online networking platforms. Another interesting fact of new media technology is that it allows users to participate in socially or politically mass discussions. In fact, these technologies provide the user a personal space to comment their opinion on the computer-generated public sphere. New media technologies make it possible for users to generate online content with a carefree attitude.

Preliminary studies on new media technologies primarily focused on limited effects of the modern means of communication (9, 10). In the field of media, for instance, barely a newspaper reporter will be able to understand the significance of blogs; hardly a television producer will be able to know why people moved from network television to online television, and so forth. One of the recent studies on New Media technologies traced the advancements in transforming knowledge in the digital format (11). It was found that users looking for information, irrespective of the place, time, and convenience led to the digital transformation through the mobile-first strategy. Other interesting findings in this field are as follows: (a) young people spend more time on social media platforms to consume information; (b) people began to verify information through the internet; (c) businesses began to promote their brands, products, and services through online social networking platforms and other social media forms; (d) People working in media began using new media to search for a new audience, cross-check the information, promote the coverage of news, and verify the accountability of the news; (e) advertising/marketing channels use new media technology to create branding for their products and services; and (f) academicians also began using different new media technologies for educational purposes (12). Recent studies underline that there is an increase in the percentage of people using various types of new media technologies from 4% in 2008 to 36% in 2018 (13, 14).

For several years, researchers have been focusing on the role of new media technologies in an attempt to understand their impact on the Digital Age. It has been traced that new media technologies build content with the help of the computer-generated public sphere. These technologies are considered an effective tool to maintain engagement in this digital age. Immediacy and cost-effectiveness are the added advantages of such technologies. As an example, one of the recent studies provided evidence of people going live on the Facebook application, which helped them to reach a much larger audience than expected (15). Although there exists no way to measure the engagement of new media, it has been emphasized that the involvement of people, the connectivity aspect between people, and their level of interest in taking participation will help the researcher define and measure the mediating effect of engagement in this digital age. With the help of advanced technologies such as web analytics, organizations from different arenas can monitor the role of people's engagement in those activities filled on new media technologies like online social networking sites, blogs, wikis, and other social media platforms. Counting the number of likes, shares, views, and comments, one can monitor the role of engagement. Processing of healthcare and travel data using machine learning algorithms in place of the traditional healthcare system to identify COVID infected persons (33).

Overall, the aim of this research paper is to investigate the mediating role of new media engagement in this digital age. Furthermore, this study defines the concept of new media, highlights the characteristics of new media technologies, and analyzes the application of new media technologies in this digital age. This study also attempts to provide a framework representing the relationship between new media and engagement in the digital age. As we all know, the twenty-first century is ruled by new media, and therefore its influence on individuals, organizations, disciplines, and so forth is relatively higher than the old media. The present study is structured as given below: Section Introduction provides an introduction to the topic; Section Literature Review examines the existing literature review on the concept of new media, its characteristics, and the impact of new media on individuals, organizations, and other areas. This section also examines the application of new media technologies in various disciplines. Section Methodology presents the methodology adopted for the study. Section Data Analysis discusses the findings and results of the study. Lastly, Section Results and Discussion discusses the conclusions derived from the study.

LITERATURE REVIEW

Understanding New Media

New Media is a new term that is universally applied in multiple ways. Seminal contributions have been made by past researchers on the concept of new media and its contextual use, with regard to communication and technology (7, 10, 16). These studies observed that communication practices, the technology or medium through which the message is communicated, and the social context in which the message is passed are quite important compared to other factors. That is, the message, channel, and source are important to communicate the information. The majority of the literature focused on the above three aspects of new media for a longer duration, besides other advanced practices and technologies like digitization, collaboration, and telecommunication (7, 10, 17). Few other studies defined new media as a web-related technology that uses advanced computer technologies to transmit messages to a larger group audience (18).

From the term new media technologies, it can be understood that such newly developed technologies use specific computer technologies to process the information and communicate it to the audience (19). A networked telecommunication system is used to interact between the users. Communication in the new media technology occurs when two electronic devices are used to share information supported by advanced technologies

and software. Emails, chat rooms, social media services, instant messaging, and so on are the best examples of new media technology communication. These platforms made the communication process easier, resulting in the establishment of a universal networked society. Additional advantages of new media services include allowing the people to work together irrespective of geographical differences, promoting globalization, establish worldwide socio-political dialogues, and framing sociological structures across the disciplines (20). A number of authors have also recognized that communication through new media has changed the whole process of communication. Most importantly, new media is known for convenience and smooth communication processes (21).

Most early studies, as well as current work, stated that the term new media encompasses various aspects like the message, the medium used to transmit the message, the technology used to transmit the message, the time during which the message is transmitted, and the social context in which the message is transmitted (22). Certainly, the term media refers to the technology through which the message is communicated. From the above statements, it is inferred that the medium through which the message is transmitted is more important than the content of the message.

Characteristics of New Media

This section presents a review of recent literature on the characteristics of new media. The wide range of new media technology characteristics includes—communication, collaboration, convergence, creativity, and community. In general, new media technologies are relatively concerned about one or other forms of communication. Blogs, which are known as one form of new media technology, allow the users to publish content on one's own web pages. Hence, the blog is an alternative web-technology communication method of transmitting information to large group audience. Furthermore, blog posts are not one-way communications, and thus demand the readers to interact/comment/post their reviews/opinions online. Therefore, there is a factor of engagement in the whole communication process, which is two-way and interactive in nature. In short, communication through new media can be categorized as collaborative, networked, and unidirectional (23). In this rapid networking age, the term viral is used to indicate the rapid speed of transmission, for instance, viral video, viral marketing, viral branding, and viral posts. Furthermore, the participants involved in new media technology communication play the role of both sender and receiver. They can even post their opinion/feedback with the support of a computer-mediated communication process.

It was reported in the literature that, if not all new media technology, at least a few of them facilitate collaboration through the internet. Much of the past collaboration happened through emails. Multiple documents could be shared with multiple people at the same time, which saves time (22, 24). Although there were minimal disadvantages of such email collaboration, it was one of the most effective means of new media communication. Such collaboration is available on Wikipedia, an online encyclopedia. Unlike wikis and emails, blogs have limited collaboration and can be shared among a specific group of people. Recent studies found that google docs permit users to share documents with each other (24). It also allows two users to work together on the same document. While social networking platforms like Facebook, Skype, WhatsApp, and others, users collaborate on the aspects of virtual communication and audio conversing. The majority of the social media platforms permit users to virtually connect with each other by means of conference calls and chats (25).

Another interesting characteristic of new media is that it fosters a sense of community. New media technologies like Skype, Facebook, Yahoo groups, Usenet, Webkinz, YouTube, and others incorporate a sense of community among the users by creating logins. The creativity of the users is the next surprising feature of new media technologies (24, 25). No user plays a passive role anymore, instead, they play an active role in transmitting the messages through web technologies. New media technologies enable its users to create videos, edit videos, post blogs, post reviews/comments on blog posts, post reviews for products/services offered by brands both national and international, and finally create content on wikis. People in this digital age not only read, listen to, or view the content posted online, but also edit, comment, and posts feedback. Hence it can be stated that there is a factor of engagement or interactivity within new media technologies. The new media technologies are broadly known for their convergence with media, technology, and other forms of web-related communication technologies. Apart from the above characteristics of new media technology, a few other features include digitized messages, multi-way communication, and a smooth flow of information (26).

New Media and Engagement

Over time, extensive literature has developed on examining the role of engagement in new media technologies. It was traced that new media becomes the most significant channel across the arenas to promote news, education, marketing, branding, and so on. Specifically speaking, people use the new technologies to share critical information with a large group audience. Multiple studies also discussed the use of new media and observed that people tend to spend more time on the newly developed technologies to read, listen, view, comment, and share information with one another (27, 28). These studies also found that user engagement plays an important role in transmitting the message to the target audience. Likewise, recent studies in the field traced that most people in this digital age posts activities through various social media platforms. In short, it can be stated that user participation is much important to process any kind of information (29–32). These new media technologies are changing the behavior of people, enabling them to interact online, create online communities, and become socially engaging. Thus, it can be argued that engagement of the audience/users/individuals is crucial in new media technology communication. Negligible studies have tried to examine the mediating role of new media engagement in this digital age and thus, this paper aims to examine the mediating role of new media engagement in this modern digital age. In the light of the above literature, the study proposes the following hypothesis as represented in **Figure 1**.

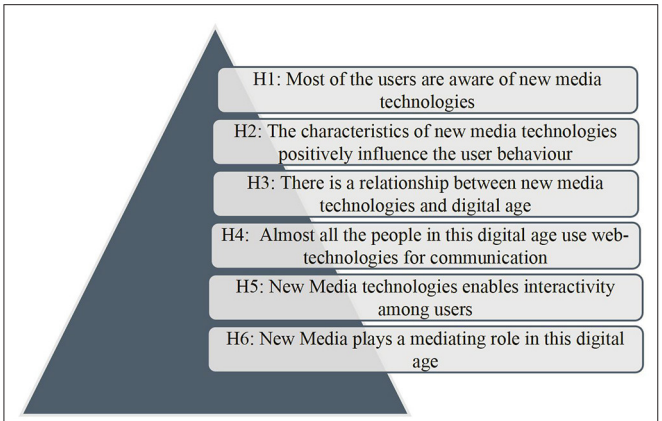


FIGURE 1 | Proposed hypotheses of the study.

METHODOLOGY

The current study mainly concentrates on exploring the mediating role of new media engagement in the digital age. It also examines the concept of new media, identifies the different types of new media technologies, and the characteristics of new media technologies. The study also examines the application of new media technology in a digital age. The analysis of the current study has been conducted through secondary literature and primary analysis as well. The study has developed hypotheses based on the analysis and review of the secondary literature. The formulated hypotheses are tested using a questionnaire. Thus, the current study gathers all the required data using a questionnaire and hence the study is descriptive. Using the results of the questionare the researcher will evaluate the hypotheses. This method involves describing the objectives, defining the population, selecting the sample, and later interpreting the data and results.

The population is nothing but the complete set of individuals or objects bound together by certain common characteristics. The population is of two types- target and accessible. The target population consists of those for whom the researcher wishes to generalize the findings of the study. While the other set of the population acts as the sub-set of it. In the current study, the total population includes all the people who fall under the age group of 18–50. From the total population, the study derives the required sample. The term sample includes the individuals selected to participate in the study. These selected participants are popularly known as respondents/subjects of research. Sampling methods are of two types—probability/random sampling and non-probability. To carry out the research, this study employs a random sampling method. At times, researchers also tend to use purposive sampling with the aim of obtaining required data from a certain area of research.

For the purpose of fulfilling the objectives of the study, the study selects a sample consisting of 100 respondents between the age group of 18–50. Both primary and secondary data are collected to find out the results. The primary data is collected using the structured questionnaire designed to collect the required data from the participants. The collected data will help the researcher to examine the mediating role of new media engagement in this digital age; review the attitude or behavior of users who use new media technologies to communicate, and examine the factor of engagement and interactivity among users with a special focus on the digital age. The researcher collects the secondary data required for the study from research articles and other journals. Besides, the study measures the formulated hypothesis with the help of 3-point Likert Scale. A pilot study is conducted to check the reliability and validity of the questionnaire, for 30 respondents. Finally, all gathered information was collected and analyzed to answer the research purpose. **Figure 2** presents the methodology followed in the study.

DATA ANALYSIS

This section of the article presents the data gathered from the selected respondents through the questionnaire. Firstly, the demographic details of the respondents are presented, along with the duration, context, and other essential details. The duration of the study lasted for 4 weeks, that is from the end of November to December. To all these 100 respondents, the questionnaire was distributed through online sites. All the participants attended the questionnaire with extreme dedication, as they were asked to fill the questionnaire in an appropriate manner. Although the questionnaire was distributed to 150 participants that include young people, employees working in various organizations like marketing, media, press, education, and self-employed; only 130 were received of which 100 were valid. These 100 valid responses were used for further analysis. **Table 1** records the demographic details of the active respondents. **Figure 3** represents the graphical representation of demographic details.

From **Table 1**, the following deductions can be recorded: (1) Most of the respondents (45%) fall under the age group of 21–30; (2) 25% of the respondents fall under the age group of 31–40; (3) 10% of the respondents fall under the age group of 41–50; (4) 10% of the respondents fall under the age group <20. Of all the active respondents, 58% of them were men and 42% of them were women, who were working in different organizations associated with brands, marketing, and so on. The majority of the respondents work in news media followed by education, marketing, students, and self-employed individuals, respectively. The statistics given in **Table 1** clearly state that almost all the participants either work in a large organization or are entrusted with new media technologies in this new digital age.

Table 2 presents the responses to the questionnaire collected from the participants of the study.

From the statistic given in **Table 2**, it can be deduced that the questionnaire was distributed to 150 people that including employees working in marketing, press, media and others, students, and self-employed individuals. Of all 150, 130 questionnaires were completed and 20 of them were incomplete. Of these 130 questionnaires, only 100 were considered valid and reliable.

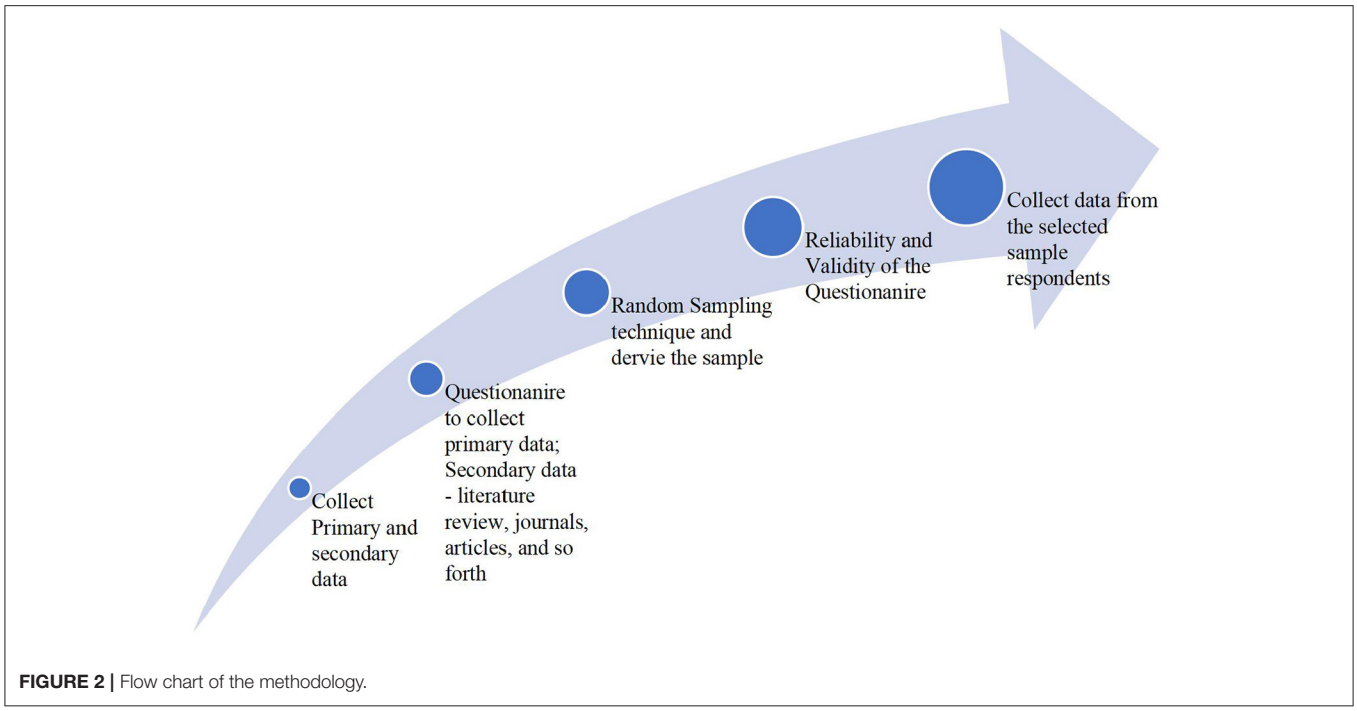


FIGURE 2 | Flow chart of the methodology.

TABLE 1 | Demographic details of the participants.

	Percentage of respondents (%)
Age	
<20	10
21–30	45
31–40	25
41–50	10
>66	0
Gender	
Male	58
Female	42
Educational qualification	
Illiterate	8
SSLC	12
Diploma	25
Graduate	55
Field	
News media	40
Marketing	20
Education	25
Students	10
Self-employed	5

RESULTS AND DISCUSSION

This section of the study presents the analysis of the data gathered through the questionnaires. The data analyzed with the help of the questionnaire are presented in **Tables 3–6** followed by a discussion of the results. An opinion poll is used to collect the data for the study. All the participants filled out the questionnaire and submitted their results online. These opinions submitted by the participants are analyzed using a 3-point Likert scale. **Table 3** records the details of how many people understand new media technology.

Understanding New Media Technology

From **Table 3**, it can be deduced that of all participants, 40% of them reported that they were fully aware of the term and concept of new media technology, 30% of them reported that they were partially aware of the term and concept of new media technology, 15% of them reported that they were unaware of term and concept of new media technology, and another 15% of them stayed neutral.

Characteristics of New Media Technologies

From **Table 4**, it can be inferred that out of 100 participants, communication features cater to 30%, convergence and community cater to 10%, the creativity of the users cater to 30%, and collaboration caters to 30%, which will positively affect the behavior of an individual/user in this digital age. This shows that almost all characteristics of new media technology impact the behavior of an individual.

New Media Technologies and Digital Age

From **Table 5**, it can be inferred that out of 100 participants 70% of them agreed that there is a positive relationship between new media technology and the digital age, while 20% of them agreed that there is a negative relationship between new media technology and the digital age. often percent stayed neutral.

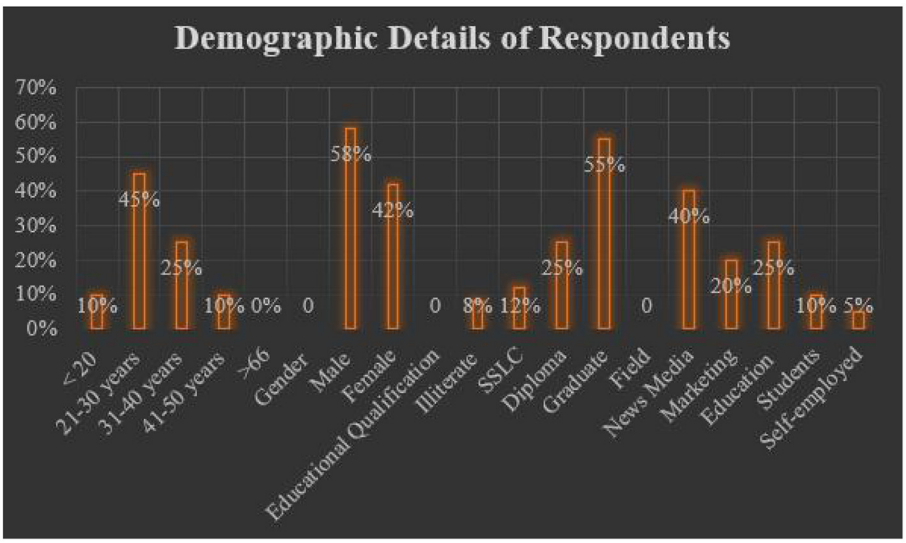


FIGURE 3 | Graphical representation of respondents' details.

TABLE 2 | Responses of the participants.

Distribution of the questionnaire	Number of respondents	Percentage (%)
Total participants	150	100
Fully completed	130	86.6
Valid questionnaires	100	66.6
Total respondents considered for this study	100	

TABLE 3 | Concept of new media technology.

About new media technology	Respondents' responses in percentage (%)
Partially aware of new media technology	30
Unaware of new media technology	15
Fully aware of new media technology	40
Neutral	15
Total	100

TABLE 4 | Characteristics of new media technologies.

Characteristics of new media technologies	Respondents' response in percentage (%)
Communication	30
Convergence and community	10
Creativity	30
Collaboration	30
Total	100

TABLE 5 | New media technology and digital age.

New media technologies and digital age	Respondents' responses by percentage (%)
Positive relationship	70
Negative relationship	20
Neutral	10
Total	100

TABLE 6 | Mediating role of engagement and interactivity.

Mediating role of engagement and interactivity	Respondents' responses by percentage (%)
New media engagement	80
New media interactivity	15
Neutral	5
Total	100

This shows that there certainly exists a relationship between new media and the digital age.

Mediating Role of Engagement and Interactivity

From **Table 6**, it can be inferred that out of 100 participants 80% of them agreed that new media plays a mediating or engaging role in this digital age to share information, view content, post feedback on blogs, and so on. Meanwhile 15% of them agreed that interactivity is essential in new media technologies, especially in this digital age, and hence interactivity plays a mediating role. Only 5% of the respondents stayed neutral. This shows that both engagement and interactivity play a mediating role between new media and the digital age. New media engagement is essential to share information

TABLE 7 | Results of rating scale.

Rating Scale	Respondents' response rate						Total	
	Always		Sometimes		Never			
	R	%	*R*	%	*R*	%	*R*	%
Most of the users are aware of new media technologies	25	25	40	40	35	35	100	100
The characteristics of new media technologies positively influence the user behavior	50	50	30	30	20	20	100	100
There is a relationship between new media technologies and digital age	40	40	30	30	30	30	100	100
Almost all the people in this digital age use web-technologies for communication	45	45	25	25	30	30	100	100
New media technologies enables interactivity among users	10	10	40	40	50	50	100	100
New media plays a mediating role in this digital age	25	25	50	50	25	25	100	100

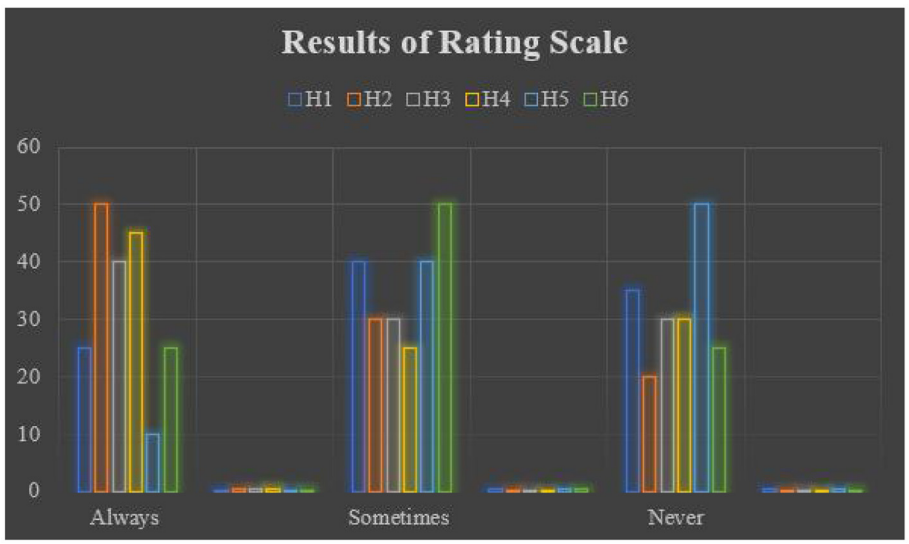


FIGURE 4 | Graphical representation of rating scale results.

with each other, develop creative content, create branding, and so on.

Results of Rating Scale

Table 7 records the statistical analysis of the rating scale selected in this study. The results of the rating scale were analyzed on a three-point Likert scale ranging from "Always" to "Never." **Figure 4** presents the graphical representation of the results of the rating scale.

The findings of the current study are given in the given **Table 8** given below.

Figure 5 shows the assessing the measures for different parameters. The current study analyzed the mediating role of engagement between new media and information in this digital age. The demographic data of respondents were also considered, which helped us to understand if the respondents were aware of the new media technologies that are rapidly spreading in this digital age. The results suggest that people in the digital age are aware of the new media web-communication technologies. Moreover, the results also indicate that almost

TABLE 8 | Item statics.

Items	Levels	
	M	MI
H1	5.23	5.29
H2	5.13	5.30
H3	4.69	5.20
H4	5.00	5.91
H5	5.19	5.90
H6	5.41	5.72

M, mean performance; MI, Mean Importance; SD, standard deviation.

90% of the respondents think that communication convergence, community, creativity, and collaboration are the five major characteristics of new media, which influences the use of new media in this twenty first century. In addition to, these characteristics are generalized and can be applied worldwide, as everything has become online these days. Furthermore, the study

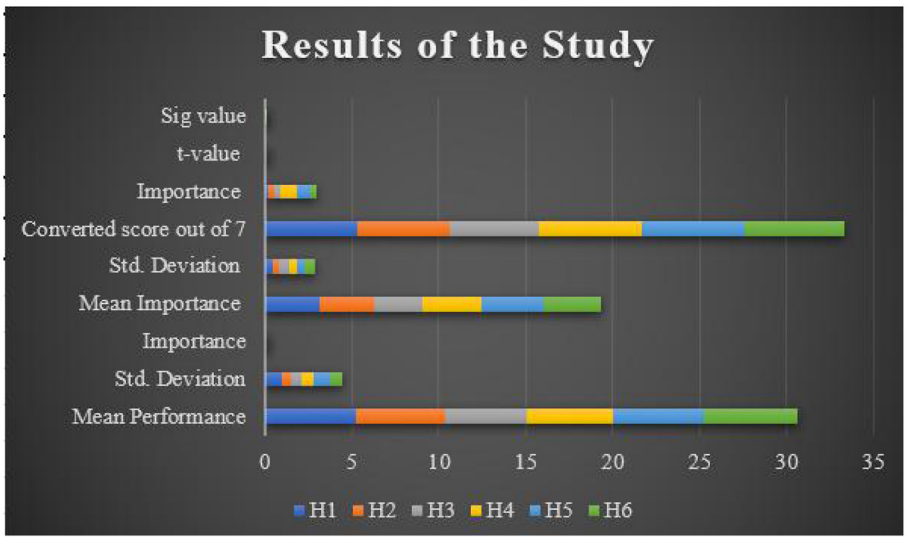


FIGURE 5 | Assessing the measures.

TABLE 9 | Assessing the measures.

Performance	Mean performance	Std. deviation	Importance	Mean importance	Std. deviation	Converted score out of 7	Importance	*t*-value	Sig value
H1	5.23	0.98	H1	3.14	0.41	5.29	0.20	−2.785	0.015
H2	5.13	0.50	H2	3.12	0.38	5.30	0.3	−3.214	0.014
H3	4.69	0.60	H3	2.82	0.59	5.20	0.39	−3.764	0
H4	5.00	0.69	H4	3.41	0.42	5.91	0.91	−12.512	0
H5	5.19	0.99	H5	3.52	0.51	5.90	0.85	−10.792	0
H6	5.41	0.71	H6	3.35	0.61	5.72	0.3	−2.412	0.022

also found that there is a positive relationship between new media technology and the digital age. A digital age permits the users to adopt different new media web-communication technologies in their daily life for easier and smooth communication. Results of the study stated that new media plays a mediating or engaging role in this digital age to share information, view content, post feedback on blogs and so on. The study also found that interactivity is essential in new media technologies, especially in this digital age, and hence interactivity plays a mediating role. Both engagement and interactivity play a mediating role between new media and digital age. New media engagement is essential to share information with each other, develop creative content, create branding, and so on. The current study can be seen as a beginning and future researchers can explore mediating role of other factors in the digital age. It is also suggested to extend the number of respondents for more information on this study. User engagement plays an important role in transmitting the message to the target audience. Likewise, it is also traced that most of the people in this digital age posts activities through various social media platforms. In short, it can be stated that user participation is very important to process any kind of information. These new media technologies are changing the behavior of the people, enabling them to interact online, create online communities, and become socially engaging. Thus, it can be argued that engagement of the audience/users/individuals is crucial in new media technology communication. **Table 9** shows the performance measure as shown below.

CONCLUSION

The present article has proposed and tested the mediating role of new media engagement in this digital age. Interactivity is essential in new media technologies, especially in this digital age, and hence interactivity plays a mediating role. Both engagement and interactivity play a mediating role between new media and the digital age. New media engagement is essential to share information with each other, develop creative content, create branding, and so on. This research analyzed the characteristics of new media technologies which affect the communication process in all organizations across various fields. The results of

the study indicated that new media technologies influence the behavior of the people, enabling them to interact online, create online communities, and become socially engaging. Thus, it can be argued that engagement of the audience/users/individuals is crucial in new media technology communication. Future researchers can adopt a scientific approach to understand the mediating role of new media engagement in this digital age. They can focus on the element of interactivity by conducting case studies to have much more appropriate results that can apply across the fields.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Wu Y. Social media engagement in the digital age: accountability or threats. *Newsp Res J.* (2018) 39:287–96. doi: 10.1177/0739532918796236
2. Molyneux L, Holton A, Lewis SC. How journalists engage in branding on Twitter: individual, organizational, and institutional levels. *Inform Commun Soc.* (2018) 21:1386–401. doi: 10.1080/1369118X.2017.1314532
3. Lysak S, Cremedas M, Wolf J. Facebook and Twitter in the newsroom: how and why local television news is getting social with viewers? *Electron News.* (2012) 6:187–207. doi: 10.1177/1931243112466095
4. Andrus DC. The wiki and the blog: toward a complex adaptive intelligence community. *Stud Intell.* 49 (2005).
5. Boulos MNK, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Med Educ.* (2006) 6:1–8. doi: 10.1186/1472-6920-6-41
6. Dearstyne BW. Blogs, mashups, & wikis: oh, my!. *Inform Manage.* (2007) 41:25.
7. Efimova L, De Moor A. Beyond personal webpublishing: an exploratory study of conversational blogging practices. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences.* Washington, DC (2005). p. 107a.
8. Friedman HH, Lopez-Pumarejo T, Friedman LW. A new kind of marketing: creating micro-niches using resonance marketing. *J Intern Commerce.* (2008) 6:83–99. doi: 10.1300/J179v06n01_05
9. Gillin P. *The New Influencers: A Marketer's Guide to the New Social Media.* Linden Publishing (2007).
10. Friedman LW, Friedman HH. *The New Media Technologies: Overview and Research Framework.* (2008). Available online at: https://ssrn.com/abstract=1116771 (accessed April 5, 2008).
11. Flew T. *New Media: An Introduction.* Oxford: Oxford University Press (2007).
12. Chester J. *Digital Destiny: New Media and the Future of Democracy.* New York, NY: New Press (2007).
13. Tsortanidou X, Daradoumis T, Barberá E. Connecting moments of creativity, computational thinking, collaboration and new media literacy skills. *Inform Learn Sci.* (2019) 120:702–22. doi: 10.1108/ILS-05-2019-0042
14. Fischer G, Lundin J, Lindberg JOJ. Rethinking and reinventing learning, education and collaboration in the digital age-from creating technologies to transforming cultures. *Int J Inform Learn Technol.* (2020) 37:241–52. doi: 10.1108/IJILT-04-2020-0051
15. Bouchard C. The United Nations in the Digital Age: harnessing the power of new digital information and communication technologies. In: editor Zaiotti RB. *Digital Diplomacy and International Organisations.* Routledge (2020). p. 101–26. doi: 10.4324/9781003032724-7
16. Zaheer L. New media technologies and youth in Pakistan. *J Res Soc Pak.* 55 (2018).
17. Adjin-Tettey TD, Mbatha B. Use and appropriation of new media technologies by teens:≪The E-teen model≫ perspective. *Galact Media.* (2019) 1. doi: 10.24411/2658-7734-2019-00001
18. Brownlee BJ. *Mediating the Syria Revolt: How New Media Technologies Change the Development of Social Movements and Conflicts.* Taylor & Francis (2018).
19. Romele A. *Digital Hermeneutics: Philosophical Investigations in New Media and Technologies.* New York, NY: Routledge (2019). doi: 10.4324/9780429331893
20. Ytre-Arne B, Das R. An agenda in the interest of audiences: facing the challenges of intrusive media technologies. *Telev New Media.* (2019) 20:184–98. doi: 10.1177/1527476418759604
21. Tomin VV, Erofeeva NE, Borzova TV, Lisitzina TB, Rubanik VE, Aliyev HK, et al. Internet media as component of information and communication environment in electoral process: features and tools. *Online J Commun Media Technol.* (2020) 10:e202011. doi: 10.29333/ojcmt/7932
22. Molyneux L, Lewis SC, Holton AE. Media work, identity, and the motivations that shape branding practices among journalists: an explanatory framework. *New Media Soc.* (2019) 21:836–55. doi: 10.1177/1461444818809392
23. Khajeheian D, Ebrahimi P. Media branding and value co-creation: effect of user participation in social media of newsmedia on attitudinal and behavioural loyalty. *Eur J Int Manage.* (2021) 16:499–528. doi: 10.1504/EJIM.2021.117524
24. Grzesiak M. *Personal Brand Creation in the Digital Age: Theory, Research and Practice.* Springer (2018). doi: 10.1007/978-3-319-69697-3
25. Kohli MD, Daye D, Towbin AJ, Kotsenas AL, Heilbrun ME. Social media tools for department and practice communication and branding in the digital age. *Radiographics.* (2018) 38:1773–85. doi: 10.1148/rg.2018180090
26. Noimor S, Chaianandech N. View of marketing in the digital age Branding, entrepreneurial innovation. *J Administr Soc Sci Rev.* (2021) 4.
27. Olsen LE. Chapter 5: future of branding in the digital age. In: *At the Forefront, Looking Ahead.* Ed. Sasson A. Universitetsforlaget (2008). p. 73. doi: 10.18261/9788215031583-2018-06
28. Ducros Passebois J, Martinez C, Euzéby F. *Branding the Performing Arts in the Digital Age: Lessons From the Opéra de Paris.* HAL (2019).
29. Li Z. Treatment and technology of domestic sewage for improvement of rural environment in China-Jiangsu: a research. *SPR.* (2021) 1:355–64. doi: 10.52152/spr/2021.154
30. Dani R, Juyal D, Rawal YS. A critical analysis of the restaurant industry's effect on environment sustainability. *SPR.* (2021) 1:435–42. doi: 10.52152/spr/2021.165
31. Salter C. Disturbance, translation, enculturation: necessary research in new media, technology, and the senses. *Vis Anthropol Rev.* (2018) 34:87–97. doi: 10.1111/var.12156
32. Alivi MA, Ghazali AHA, Tamam E, Osman MN. A review of new media in Malaysia: Issues affecting society. *Int J Acad Res Bus Soc Sci.* (2018) 8:12–29. doi: 10.6007/IJARBSS/v8-i2/3849
33. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Publ Health.* (2020) 8:357. doi: 10.3389/fpubh.2020.00357

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yujie, Al Imran Yasin, Alsagoff and Hoon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 04 May 2022
doi: 10.3389/fpubh.2022.861074

Financial Budgets of Technology-Based SMEs From the Perspective of Sustainability and Big Data

Guobiao Zhang[1] *and Tao Wang*[2]*

[1] School of Business, Zhejiang University City College, Hangzhou, China, [2] School of Business, Zhejiang University City College, Hangzhou, China

In this contemporary world, the words data and sustainability play a crucial role in determining the financial budgets of small and medium-sized enterprises (SMEs). Usually, it is stated that the survival of small and medium-sized enterprises (SMEs) is directly proportional to the growth and sustainability factor of the nation. The economic sustainability of a nation is dependent on appropriate functioning of SMEs. Any kind of direct impact on the working of SMEs will have its impact on the whole economy of a nation. There are different factors such as lack of financial capacity, low market demands, restrictions with regard to the capital, and barriers in the supply chain that affect the sustainability of SMEs worldwide. Nevertheless, small, and medium sized enterprises around the world are greatly investing on skills, innovation, and other capital related resources to mark up the demands of the external market. The main objective of this study is to examine the financial budgets of technology-based SMEs from the perspective of sustainability and big data. For this, the study collects data through a questionnaire from 1,800 Small and Medium Sized Enterprises. Based on a detailed and careful examination of the data, only 1,400 of the responses received were considered valid (79.75%). To test the hypothesis stated, the study employs structural equation modeling. This will help the researcher to examine the direct effect of financial budget and technology adaption of SMEs from the perspective of sustainability and big data. Results of the study stated that SMEs sustainability and big data are directly and positively related to the financial budget planning of technology-based SMEs. The study also found that big data plays an important role in the businesses, specifically for their own growth.

Keywords: small and medium sized enterprises, big data, sustainability, finance, financial budgets, technology adaptation

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Chennareddy Vijay Simha Reddy,
Middlesex University, United Kingdom
Praveen Kumar,
VIT University, India
Rutvij H. Jhaveri,
Pandit Deendayal Petroleum University, India

***Correspondence:**
Tao Wang
twang2022@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *24 January 2022*
Accepted: *23 February 2022*
Published: *04 May 2022*

Citation:
Zhang G and Wang T (2022) Financial Budgets of Technology-Based SMEs From the Perspective of Sustainability and Big Data. Front. Public Health 10:861074. doi: 10.3389/fpubh.2022.861074

INTRODUCTION

Recently, businesses across the world are facing extreme sustainability issues and thereby adversely impacting the economy of the nations. Due to the current societal circumstances, most of the businesses are enforced to formulate new ideas to deal with the problems associated with the pressurized societal situation. Especially, Small and Medium sized Enterprises (SMEs) recorded the most pressing difficulties including financial and Non-financial. There are other factors like limited market demand, supply chain disruption, and trade discontinuation that will impact the

economy of the nation (1). The current environmental situations in the world critically affected the economy of the nation, nevertheless, the most affected sector was identified as Small and Medium sized Enterprises (SMEs) (2).

Very recently, (3) provided evidence stating that SMEs act as the major driving force in increasing the growth of the nation. SMEs also create employment opportunities, open markets for economic sustainability, and develop trade liberalization. Because of financial and economic crisis, several small and medium sized enterprises are facing challenges. Most of the SMEs are failing to contribute the economic development of their nation. Several studies, for instance (4, 5), have revealed that SMEs are the gravest hit industries, especially during any type of crisis. It was noted that these SMEs knowledge level was low, susceptible, less reliant, and so on. Moreover, SMEs were greatly dependent on financial agents who support them during a crisis (6). Several studies also recorded that SMEs are facing multiple challenges with regard to financial obligation such as paying credit to the financial institutions, identifying shortages in the inventory, and identifying shortage in the operations.

To achieve the goal of sustainability, it becomes equally important for the SMEs to predominantly survive and contribute to the total economy of the country. Although there exist several ways that hinder the growth of SMEs, it becomes crucial for an SME to survive by itself and contribute to the whole economy of the country (7). Despite SMEs have sorted their own methods/ways to deal with survival, it had affected their businesses in a positive and negative manner. Usually, owners of the businesses are concerned about the operations of the SMEs, which included financial budget activities, financial disruption activities, creating unemployment, creating social imbalance, creating security imbalance, and planning the whole production. It was also observed that these operations had had a deep impact on the society (8).

Preliminary studies of SMEs sustainability primarily focused on exploring factors that affected the growth and development of small and medium sized enterprises. Factors such as adopting technology in the SMEs, implementing financial budgets, appropriating financial planning, effective big data management, and so forth affected the growth and functioning of SMEs (9). A growing number of literatures has proved that technology-based SMEs perform well in terms of dealing with sustainability and big data. It was also found that there is a positive relationship between adapting technology and sustainability of SMEs. From the above statement, it can be deduced that the technology-based functioning SMEs find advantages in the market and thus moving toward the path of sustainability (10). Although there exist several studies exploring the relationship between finance and the growth of SMEs, so far, no study has tried to analyze the financial budgets of technology-based SMEs from the perspective of sustainability and big data.

Past studies confirmed that data is regarded as one of valued assets for any business, and it's been increasing rapidly day-by-day. This valued asset is popularly termed as big data that comes with 3 peculiar features. These three features are identified as having high volumes of data, variety of data types, and velocity to process the data. Almost all the organizations in the world are striving hard to discover all the possible ways to utilize this big data and develop their business. Majority of the studies confirmed that big data is quite important in the growth and development of a business (11). It was also added that, both large and small organizations are focusing on accurate and valuable data to make some important decisions related to business development. Likewise, few other studies demonstrated that big data supports SMEs to reach out to their target audience, understand their preferences, identify their needs, and then meet them in new ways (9, 12). Hence, it can be understood that big data plays a significant role in the growth of an SME, which is why SMEs should adopt big data in their business. Therefore, this study focuses on examining the financial budgets of technology-based SMEs from the perspective of sustainability and big data, which are considered as the backbone of a nations' economy. The other objectives of the study are to improve the

TABLE 1 | Research details.

Headers	Comments
Sector	Small and Medium sized Enterprises (SMEs)
Location	China
Methodology	Questionnaire–Structured in nature
Sampling Technique	Technique of Random Sampling
Selected respondents	1,800 (1,400)
Period of data collection	November 2021 to December 2021

TABLE 2 | Variables of the study.

Name of the Variable	Variable definition	Items (5-point likert scale)
SMEs sustainability	Enhance services to customers, improve supply chain linkage, and strengthen their knowledge of market and trade relationships	4
Adopting innovative technology within SMEs	Adopting innovative technological advancements within the small and medium sized enterprises and large multinational organizations will help them to enhance their services to customers, improve supply chain linkage, and strengthen their knowledge of market and trade relationships.	5
Innovative finance	Innovative ways to manage financial products and services for effective operation of a business.	5
Big data	Big Data refers to massive volumes of data, which expands the business with new opportunities and challenges	5

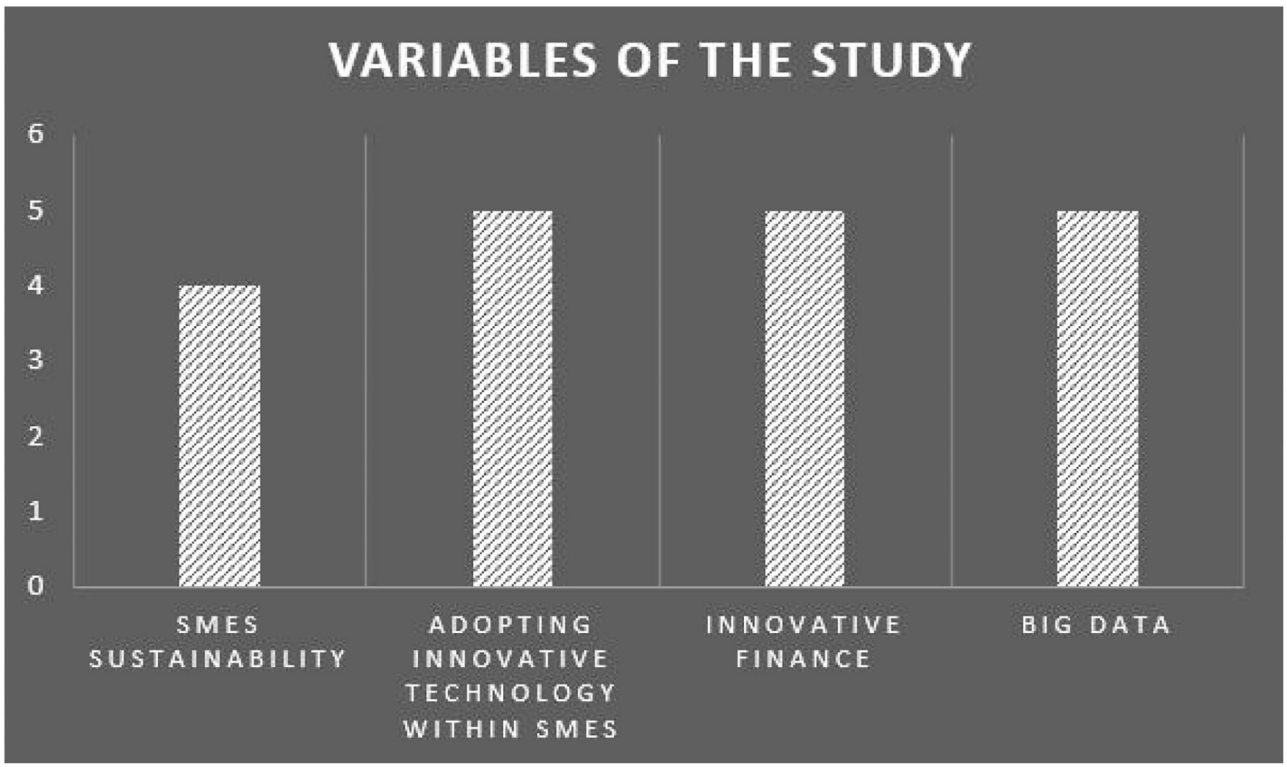


FIGURE 1 | Graphical representation of the identified variables in the study.

business process of SMEs by adopting new technologies and planning for financial budgets, which will effectively promote the growth of SMEs. This study will address the existing research gap by examining the process of planning financial budgets to the technology-based SMEs through the perspective of big data and sustainability. It will also examine the relationship between finance and SMEs; technology and SMEs; big data and SMEs; and try to understand this complex relationship from the perspective of SME sustainability.

LITERATURE REVIEW

Relationship Between Finance and SMEs

The growth of small and medium sized enterprises is directly related to the aspect of finance. Innovations with regard to finance, especially in the smooth functioning of SMEs play a critical role in various forms. In order to perform well, external and internal financial resources are necessary within an SME (13). Generally speaking, the performance of a firm is predominantly determined by intellectual resources and other major decisions. Moreover, any firm seems to face external and internal constraints (14). Past studies advocated that SMEs receive financial support from different traditional sources like merchant banks, development banks, commercial banks and so on. Most of the existing financial institutions were established to maintain the financial budgets of private organizations. For a smooth flow of business, these banks serve them with their effective channels (15). To manage the business in a much effective manner, every organization need loans from bank. From the above it can be understood that there is a direct relationship between credit of banks i.e.,' bank loans and the performance of a company, be it large or small and medium sized enterprises. The economic activity between a bank and an SME, is thus correlated.

Recent studies discussed about the effectiveness of the integration of e-commerce within Small and Medium sized Enterprises and found that latest technologies like mobile banking permits many SMEs the flexibility to settle down their financial transactions through the application (16). These new technologies also help the businesses to effectively perform financial transactions and reduce costs. It was also found that adopting mobile banking services within the small and medium sized enterprises will improve the diversity of business, reduce the cost, maintain the security, provide settlements with regard to finance. To perform monetary transactions, mobile commerce technology uses computer networks and other wireless networks. Previous studies observed that there has been a steady increase in the number of people using mobile applications and devices to buy products/services online. Using smartphone technology within the businesses can reduce the normal risks associated with finance. Hence, the study proposes the following hypothesis:

H1: There is a positive relationship between financing and SMEs sustainability.

Relationship Between Technology and SMEs

Previous studies have shown that technology plays a crucial role in the effective functioning of an enterprise. Adapting to the innovative technology will permit the business entity to establish their brand in the market. It will also boost the competitor factor within the SMEs in the global market (17). As a result of globalization, the economic and financial condition of the SMEs are continuously persisting in terms of attaining large economy scale. Moreover, the revolution created by technology has led to several problems addressing the SMEs and the owners of SMEs had to respond to those issues. Interestingly, it was found that innovations in technology enhanced the business of SMEs and as a result SMEs witnessed a long-term growth. Hence, from this perspective it becomes important to understand the role of technology in achieving SMEs sustainability (18, 19).

Few other studies stressed that certain SMEs export their products/services at international level. They broaden their market with the help of innovative technologies. Technology based SMEs focused on pushing themselves to enter international market and expand their potential. When the SMEs are exposed to the overseas market, equivalently they et an opportunity to expand themselves and efficiently communicate, transport, and trade their services (20–22). Above all, advanced technologies enhanced the internal and external communication process from the organization perspective. It also boosted the competitiveness among SMEs by improving the quality of goods and services. With the advancement of technology, supply chain linkages has also improved by aiding new services.

Another important aspect of using technology in SMEs is that it gives excellent outcome in the market. Use of advanced technology strengthen the condition of firms in the markets at international and national level (23). Recent studies found the strong appeal on adapting technology within the SMEs. For various reasons like social culture, technology adaptation, economic structure, and political condition of nation, sometimes the business can be prolonged. Numerous studies discussed the advantages of implementing technologies in SMEs (24–26). Nevertheless, few other studies aimed to focus on adopting technology within small and medium sized enterprises. Both developing and developed countries are interested in adopting technology to the SMEs, which will enhance the performance of firms. Innovations in technology will enhance the firm performance, thereby making SMEs sustainable. Hence, it can be commented that technology-based SMEs are critical in developing the economy of nation. Thus, the study proposes the following hypothesis:

H2: There is a positive relationship between adopting technology and SMEs sustainability.

Big Data and SMEs

There exists a considerable body of literature on the importance of data within the companies. Data is regarded as an important asset of SMEs. SMEs across the world are exploring different ways to tackle the data, which is usually called as big data. Big data is not only used in large multinational organizations, but also in SMEs. With the aim of improving the quality of a business, SMEs make use of big data to make accurate and quick decision in the need of hour (27, 28). The literature review shows that the use of big data in the business will enhance the whole process of business, and for this reason, SMEs are thinking on adopting big data to businesses (29–31). The use of big data technologies will help SMEs in deriving market value. The best way to deal with the challenges in the business is to adopt big data in SMEs (27). However, researchers have shifted their focus on big data, on discovering the benefits of big data.

A series of recent studies has indicated that multinational organizations are buying big data. Even startups are purchasing big data technologies to understand the new market, new clients, new business and so forth. Big data will also help SMEs to cut down the costs of business and enter the market. Transforming the big data to business applications will cost them real difficulty. Investing capital in the big data hardware and software is high.

TABLE 3 | Demographic details of the respondents.

Details		Number	Percentage %	Total count
Gender	Male	920 (M)	65.7%	
	Female	480 (F)	35.1%	
Business nature	Manufacturing	820	52.01%	
	Services	580	47.12%	
Establishment year	<10	242	18.12%	1,400
	10–15 years	561	40.10%	
	15–20 years	322	22.09%	
	>20 years	272	19.61%	
Employee count	>150	172	13.01%	1,400
	25–150	687	49.21%	
	<25	536	39.41%	
Revenue	>10 million	371	27.21%	1,400
	5–10 million	582	42.60%	
	<5 million	442	32.23%	

TABLE 4 | Validating the instrument used in the study.

	FI	AT	SME
FI 1	0.921		
FI 2	0.882		
FI 3	0.945		
FI 4	0.891		
FI 5	0.820		
AT 1		0.991	
AT 2		0.891	
AT 3		0.912	
AT 4		0.890	
SME 1			0.893
SME 2			0.882
SME 3			0.960
SME 4			0.921
SME 5			0.881

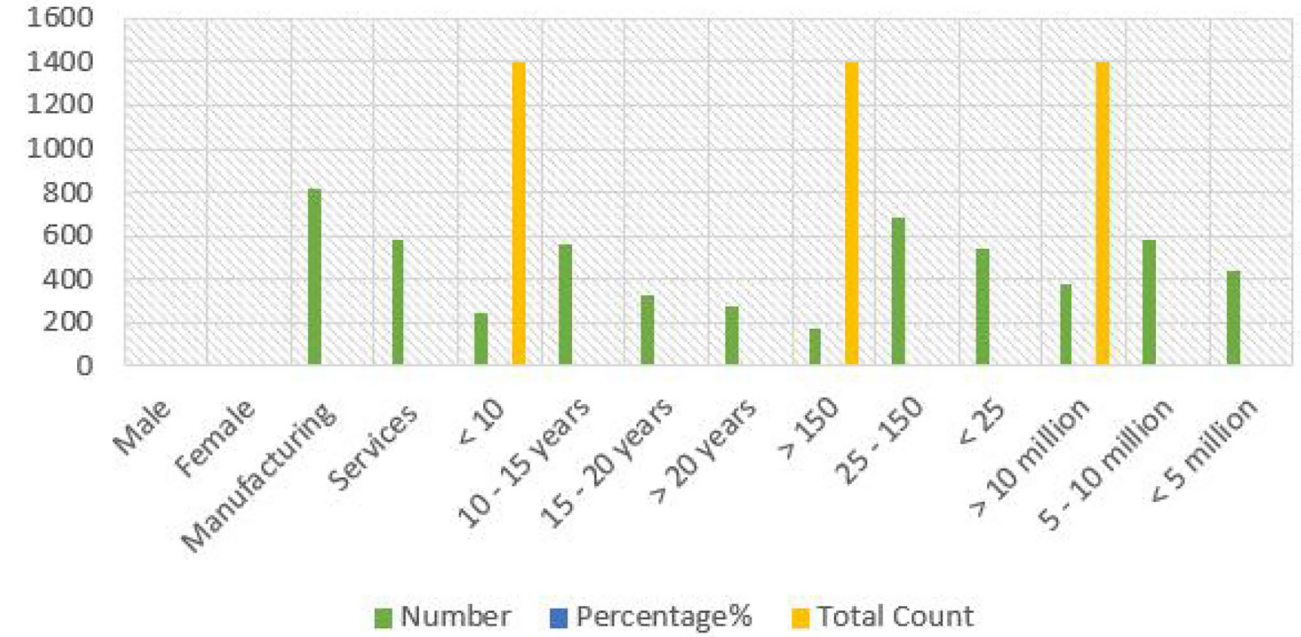


FIGURE 2 | Graphical representation of demographic details of respondents.

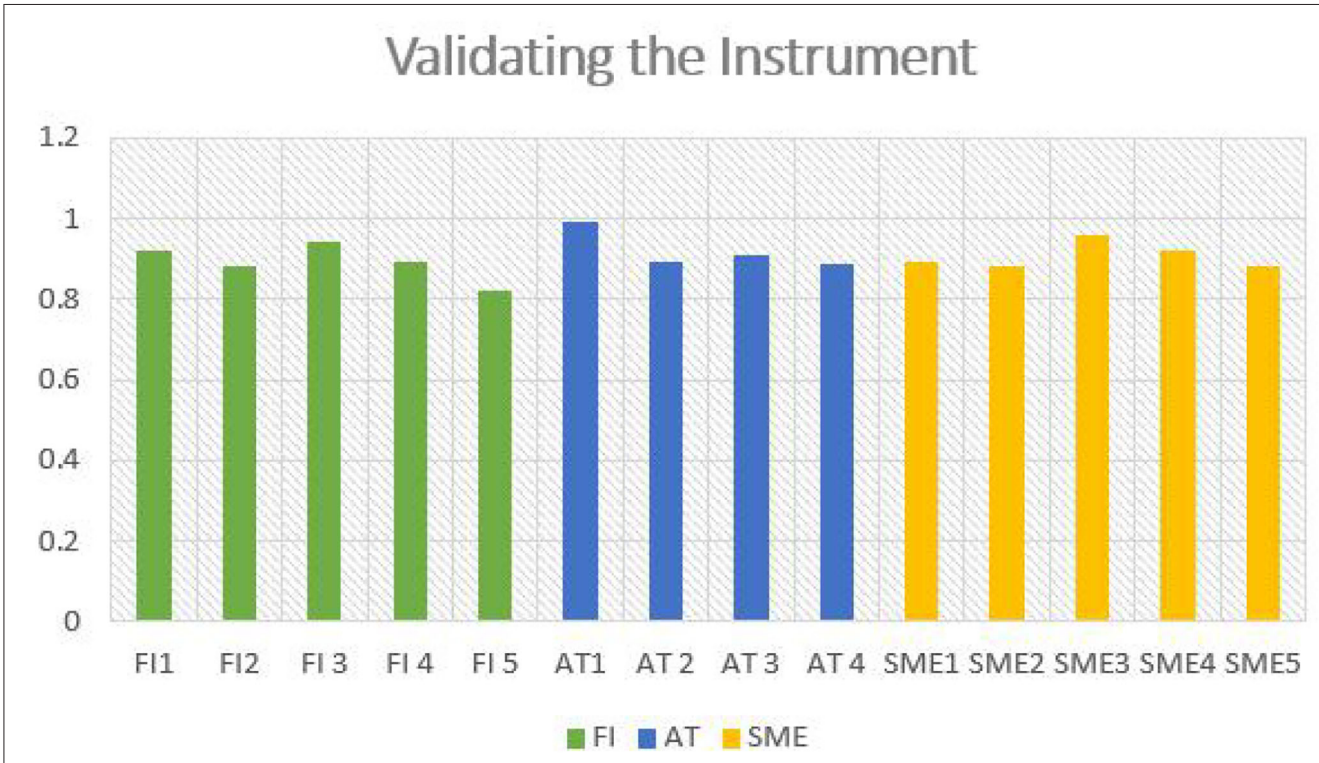


FIGURE 3 | Graphical representation of validating the instrument used in the study.

Few other studies revealed that technology adaptation, financial innovations, and big data are regarded as the strong pillars of SME (14, 16). In fact, organizations that use big data can increase their productivity and growth, which will improve the performance of enterprises. It will also help SMEs to understand the behavior of the market. Therefore, the study proposes the following hypothesis:

H3: There is a positive relationship between Big Data and SMEs sustainability.

H4: Big Data increases the growth and development of SMEs.

H5: Innovative Finance, Technology adaptation and Big Data technologies are essential in SMEs sustainability.

TABLE 5 | Results of hypothesis testing.

Relationships	Path	Std. Dev	*t* Value	*p*-value
There is a positive relationship between financing and SMEs sustainability	0.169	0.121	4.425	0.002
There is a positive relationship between adopting technology and SMEs sustainability	0.275	0.071	6.718	0
There is a positive relationship between Big Data and SMEs sustainability	0.212	0.081	5.387	0.014
Big Data increases the growth and development of SMEs	0.198	0.079	2.671	0.31
Innovative Finance, Technology adaptation and Big Data technologies are essential in SMEs sustainability	0.19	0.08	2.781	0.124

METHODOLOGY

The current study mainly concentrates on examining the financial budgets of technology-based SMEs from the perspective of sustainability and big data. It also examines the relationship between innovative finance and SMEs, adapting new technologies within the SMEs, adopting to big data technologies from the perspective of SME sustainability. The analysis of the current study has been conducted through secondary literature and primary analysis as well. The study has developed hypothesis based on the analysis and review of the secondary literature. The formulated hypothesis is tested using a structured questionnaire. Thus, the current study gathers all the required data using questionnaire and hence the study is descriptive. Using the questionnaire's results the researcher will evaluate the hypothesis. This method involves describing the objectives, define the population, select the sample, and later interpret the data and results. Population is a complete set of individuals who are bound together by certain common characteristics. Population is of two types- target and accessible. Target population consists of those whom the researcher wishes to generalize the findings of the study. In the current study the total population include 7.2 million SMEs in China. From the total population, the study derives the required sample. The term sample includes the individuals selected to participate in the study. These selected participants are popularly known as respondents/subjects of research. Sampling methods are of two types–probability/random sampling and Non-probability. To carry out the research, this study employs random sampling method. At times, researchers also tend to use purposive sampling with the aim of obtaining required data from certain area of research.

For the purpose of fulfilling the objectives of the study, the study selects the sample consisting of 5,210 SMEs located in China. This study focused on those SMEs in China that were relatively easy to connect using an electronic channel. Prior to the administration of the questionnaire prepared, consent from the SMEs was imitated as the first step. Only after the participants acknowledgment, the questionnaire link was distributed to the selected sample. Out of 5,210 sample, around 4,000 of them acknowledged and agreed to participate in the study. So, in total 4,000 SMEs sent their prior consent, out of which only 2,000 were considered valid. Based on the analysis of the secondary literature, the researcher prepared a structured questionnaire, which addressed the factors like innovations in finance, adapting new technologies within the SMEs, adopting big data within SMEs from the perspective of SME sustainability. This structure questionnaire will allow the selected respondents to share their views, opinions, beliefs, attitudes, and values on the research topic. The data was collected during the period of November 2021 to December 2021 and sample of 3,000 was considered along with their responses. **Table 1** presents the technical details of the research. In total, 1,800 responses were collected back and after the screening process, data of 1,400 respondents were finalized and considered for further analysis. Both primary and secondary data is collected to find out the results. The primary data is collected using the structured questionnaire designed to collect the required data from the participants. The collected data will help the researcher to examine the financial budgets of technology-based SMEs from the perspective of sustainability and big data. The researcher collects the secondary data required for the study from research articles and other journals. Finally, all gathered information was collected and analyzed to answer the research purpose.

The present identifies different variables like innovative finance, innovative technology, SMEs sustainability, and Big Data technology along with the measurement model. This model is presented in **Table 2** and its graphical representation is presented in **Figure 1**.

To analyze the data collected, the researcher employs SPSS and AMOS −26.0 version. To study the relationship between the variables identified, the study adopts structural equation modeling. Furthermore, the selected SMEs will be able to provide accurate and precise information on the questions covered in the questionnaires related to the research topic. As the SEM method is highly resilient and efficient in modern times, the researcher decided to use the same. This method will help the researcher to examine the relationship between the variables identified by testing the proposed hypothesis. This method also

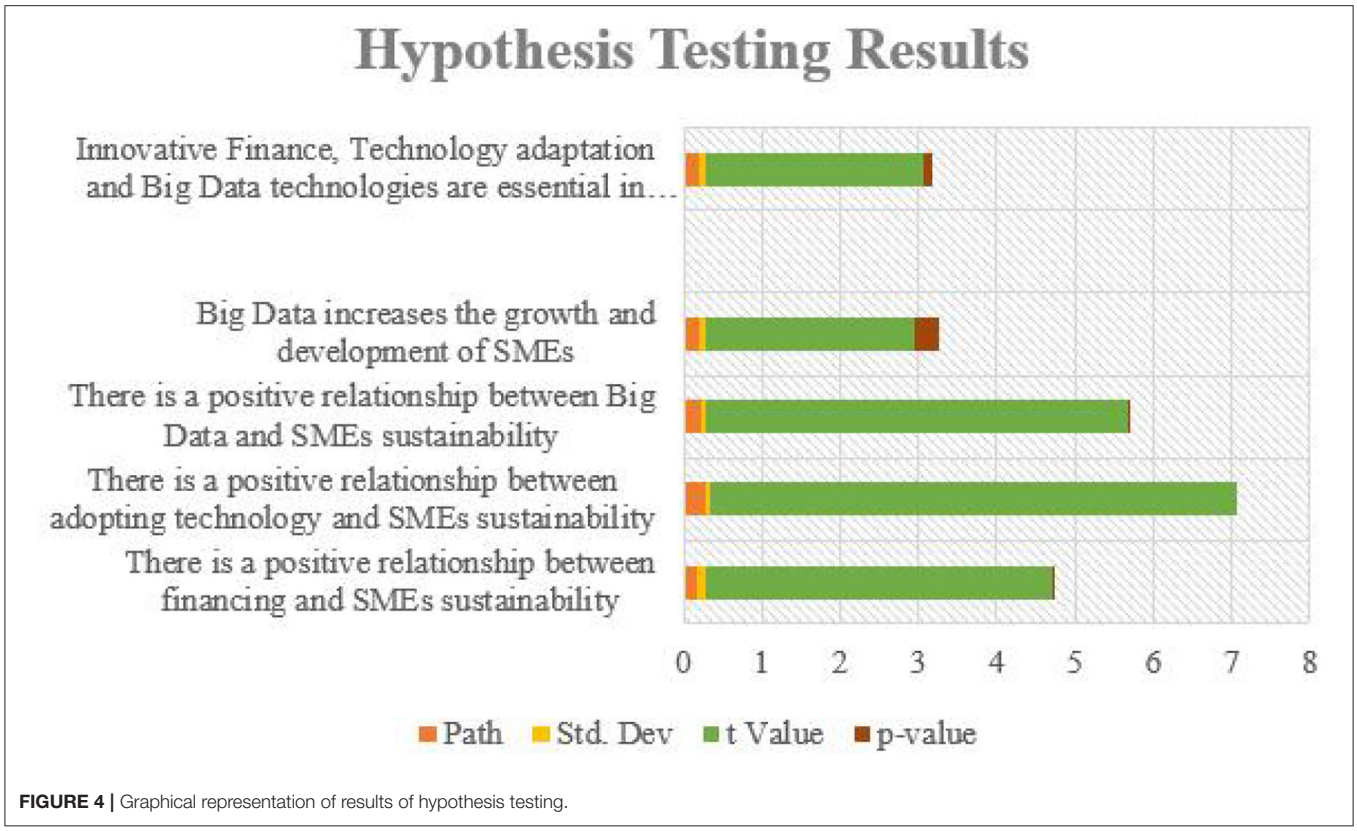


FIGURE 4 | Graphical representation of results of hypothesis testing.

focuses on the appropriateness of the results and hence was taken into consideration.

Although the owners of SMEs were reluctant to share the details of firms such as firm performance, innovations in finance budgeting, adopting innovative technologies, and adopting big data technologies, from the perspective of SME sustainability initially, later, they understood the need to evaluate the financial budgets of technology-based SMEs from the perspective of big data and sustainability and thereby opened, which helped the researcher to record the accurate data. Past researchers also tried to measure the sustainability of SMEs, which were in line with the opinions of the respondents, marked in 5-point likert scale, scaling from extremely poor to extremely good. The study also recorded responses from the participants on innovative ways to manage finance in SMEs from the perspective of sustainability and big data, and even their attitude toward the concept of innovative finance on the 5-point likert scale. Lastly, the study also recorded the responses of the respondents on adapting to innovative technologies and big data within SMEs that will enhance the sustainability on a 5-point likert scale.

RESULTS AND DISCUSSION

The current study confirmed the following findings. **Table 3** records the demographic details of the respondents, who participated in the structured questionnaires. When analyzing the data of SMEs, it was understood that 65.75 of them are males and 35.1 of them are females, spread across business units like manufacturing and services, fashion houses, units of handicraft and boutique. 52.01% of the respondents come under the manufacturing businesses and 47.12% of them belong to service units like beauty parlor, fast food, book shops and so on. Likewise, 40.10% of the SMEs were established for almost more than 10 years and come in between 15 years, 22.09% of the SMEs were established for more than 15 to 20 years, 19.61% of the SMEs were established for more than 20 years, and least 18.12% of them were established for <10 years. With regard to employee count, 49.21% of them disclosed high level of engagement in operations, 39.41% them disclosed a mediating level of employee engagement, and 13.01% exhibited low levels of employee engagement. Lastly, with regard to the employee revenue, 42.60% of them earned almost 5–10 million in a year, followed by 32.23% (<5 million) and 27.21% (<10 million).

The validation of the instrument used in the study is recorded in **Table 4**. From the table, it can be understood that the skewness test range results suggest that innovative finance, adapting technology within SMEs from the perspective of big data and sustainability are equally distributed. The extreme loading values presented in the table state that the instrument used is highly reliable and valid, which is calculated as 0.7. in terms of implementation. Moreover, the results presented in **Table 4** also confirm the absence of multicollinearity in the data gathered. **Figure 2** presents the graphical representation of validating the Instrument used in the study.

The result of the empirical study shows that innovative ways to manage finance within the SMEs will play a crucial role from the perspective of SME sustainability, as the coefficient value seems to be higher than the cut-off value. Likewise, the results associated with variable adopting technology within the SMEs also confirmed that they seem to significantly impact the smooth functioning of the SMES, especially when advanced technologies are implemented in line with SME sustainability and big data. Al the identified variables seem to efficient and hence the results can be considered as reliable. The findings of the study also disclosed that there is no reliability and validity issue with regard to the identified constructs. The **Figure 3** shows the Graphical Representation of Validating the Instrument used in the study.

In addition to, the study presented the results in measurement models including all the identified various variables. **Table 5** presented the measurement model results of the study, which indicate that there exists a relationship between the constructs–innovative finance, adopting technology within the SMES, adapting to big data technologies from the perspective of SME sustainability and big data. **Figure 4** represents the results of Hypothesis testing.

The study found that the hypothesis proposed in the study have high impact on the SME sustainability. Results of the study document that there is a positive relationship between SME and innovative finance, adapting to technological innovations, adapting to big data technologies, especially from the perspective of SME sustainability and big data. The study also found that the SME in China are high in number and hence the operations of the SMEs should focus on adopting to innovative finance, technology from the perspective of big data and sustainability, which stay significant for the survival of SMEs in this modern world. Even big data technologies will create new challenges and opportunities to the SMEs in contemporary world. Furthermore, there is a positive relationship between adopting technology and SMEs sustainability, which will expose SMEs to new opportunities and challenges. In short, it can be summarized that Innovative Finance, Technology adaptation and Big Data technologies are essential in SMEs sustainability.

CONCLUSION

The present article has proposed and tested the relationship between Innovative Finance, Technology adaptation and Big Data technologies, which play an essential role in SMEs sustainability. Hence this study examined the financial budgets of technology-based SMEs from the perspective of sustainability and big data. The results of the study indicated that there is a positive relationship between SME and innovative finance, adapting to technological innovations, adapting to big data technologies, especially from the perspective of SME sustainability and big data. There is a positive relationship between adopting technology and SMEs sustainability, which will expose SMEs to new opportunities and challenges. Thus, it can be argued that can be summarized that Innovative Finance, Technology adaptation and Big Data technologies are essential in SMEs sustainability.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. Saidu M, Aifuwa HO. Coronavirus pandemic in nigeria: how can small and medium enterprises (SMEs) cope and flatten the curve. *Eur J Account Finance Invest*. (2020) 6:55–61. Available online at: https://ssrn.com/abstract=362148411
2. Manyati TK, Mutsau M. Exploring technological adaptation in the informal economy: a case study of innovations in small and medium enterprises (SMEs) in Zimbabwe. *Afr J Sci Technol Innov Dev*. (2019) 11:253–9. doi: 10.1080/20421338.2018.1552650
3. McGuinness G, Hogan T, Powell R. European trade credit use and SME survival. *J Corp Finance*. (2018) 49:81–103. doi: 10.1016/j.jcorpfin.2017.12.005
4. Das S, Kundu A, Bhattacharya A. Technology adaptation and survival of SMEs: a longitudinal study of developing countries. *Technol Innov Manage Rev*. (2020) 10:64–72.. doi: 10.22215/timreview/1369
5. Okpara JO. Factors constraining the growth and survival of SMEs in Nigeria: implications for poverty alleviation. *Manage Res Rev*. (2011) 34:156–71. doi: 10.1108/01409171111102786
6. Sui S, Baum M. Internationalization strategy, firm resources and the survival of SMEs in the export market. *J Int Bus Stud*. (2014) 45:821–41. doi: 10.1057/jibs.2014.11
7. Bourletidis K, Triantafyllopoulos Y. SMEs survival in time of crisis: strategies, tactics and commercial success stories. *Procedia-Soc Behav Sci*. (2014) 148:639–44. doi: 10.1016/j.sbspro.2014.07.092
8. Hassan MU, Iqbal Z, Malik M, Ahmad MI. Exploring the role of technological developments and open innovation in the survival of SMEs: an empirical study of Pakistan. *Int J Bus Forecast Mark Intell*. (2018) 4:64–85. doi: 10.1504/IJBFMI.2018.088629
9. Shin K, Park G, Choi JY, Choy M. Factors affecting the survival of SMEs: a study of biotechnology firms in South Korea. *Sustainability*. (2017) 9:108. doi: 10.3390/su9010108
10. Naradda Gamage SK, Ekanayake EMS, Abeyrathne GAKNJ, Prasanna RPIR, Jayasundara JMSB, Rajapakshe PSK. A review of global challenges and survival strategies of small and medium enterprises (SMEs). *Economies*. (2020) 8:79. doi: 10.3390/economies8040079
11. Shah HA, Yasir M, Majid A, Javed A. Impact of networking capability on organizational survival of SMEs: mediating role of strategic renewal. *Pak J Commer Soc Sci*. (2019) 13:559–80. Available online at: https://www.econstor.eu/handle/10419/20526724
12. Kang JW, Heshmati A. Effect of credit guarantee policy on survival and performance of SMEs in Republic of Korea. *Small Bus Econ*. (2008) 31:445–62. doi: 10.1007/s11187-007-9049-y
13. Pu G, Qamruzzaman M, Mehta AM, Naqvi FN, Karim S. Innovative finance, technological adaptation and SMEs sustainability: the mediating role of government support during COVID-19 pandemic. *Sustainability*. (2021) 13:9218. doi: 10.3390/su13169218
14. Prasanna RPIR, Jayasundara JMSB, Naradda Gamage SK, Ekanayake EMS, Rajapakshe PSK, Abeyrathne GAKNJ. Sustainability of smes in

the competition: a systemic review on technological challenges and sme performance. *J Open Innov*. (2019) 5:100. doi: 10.3390/joitmc5040100
15. Klewitz J, Hansen EG. Sustainability-oriented innovation of SMEs: a systematic review. *J Clean Product*. (2014) 65:57–75. doi: 10.1016/j.jclepro.2013.07.017
16. Chege SM, Wang D. The influence of technology innovation on SME performance through environmental sustainability practices in Kenya. *Technol Soc*. (2020) 60:101210. doi: 10.1016/j.techsoc.2019.101210
17. Bos-Brouwers HEJ. Corporate sustainability and innovation in SMEs: evidence of themes and activities in practice. *Bus Strategy Environ*. (2010) 19:417–35. doi: 10.1002/bse.652
18. Biondi V, Iraldo F, Meredith S. Achieving sustainability through environmental innovation: the role of SMEs. *Int J Technol Manag*. (2002) 24:612–26. doi: 10.1504/IJTM.2002.003074
19. Shihadeh F, Naradda Gamage SK, Hannoon A. The causal relationship between SME sustainability and banks' risk. *Econ Res*. (2019) 32:2743–60. doi: 10.1080/1331677X.2019.1655465
20. de Sousa Jabbour ABL, Ndubisi NO, Seles BMRP. Sustainable development in Asian manufacturing SMEs: progress and directions. *Int J Prod Econ*. (2020) 225:107567. doi: 10.1016/j.ijpe.2019.107567
21. Fuller-Love N. Management development in small firms. *Int J Manage Rev*. (2006) 8:175–90. doi: 10.1111/j.1468-2370.2006.00125.x
22. Munjal S, Requejo I, Kundu SK. Offshore outsourcing and firm performance: moderating effects of size, growth and slack resources. *J Bus Res*. (2019) 103:484–94. doi: 10.1016/j.jbusres.2018.01.014
23. Li Z. Treatment and technology of domestic sewage for improvement of rural environment in China-Jiangsu: a research, *SPR*. (2021) 1:355–64. doi: 10.52152/spr/2021.154
24. Dani R, Juyal D, Rawal YS. A critical analysis of the restaurant industry's effect on environment sustainability. *SPR*. (2021) 1:435–42. doi: 10.52152/spr/2021.159
25. Begenau J, Farboodi M, Veldkamp L. Big data in finance and the growth of large firms. *J Monet Econ*. (2018) 97:71–87. doi: 10.1016/j.jmoneco.2018.05.013
26. Gunasekaran A, Yusuf YY, Adeleye EO, Papadopoulos T. Agile manufacturing practices: the role of big data and business analytics with multiple case studies. *Int J Prod Res*. (2018) 56:385–97. doi: 10.1080/00207543.2017.1395488
27. Iqbal M, Kazmi SHA, Manzoor A, Soomrani AR, Butt SH, Shaikh KA. A study of big data for business growth in SMEs: opportunities & challenges. In: 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1–7). IEEE. (2018)
28. Ogrean C. Relevance of big data for business and management. Exploratory insights (Part I). *Stud Bus Econ*. (2018) 13:153–63. doi: 10.2478/sbe-2018-0027
29. Mikalef P, Pappas I, Krogstie J, Pavlou PA. Editors. Big Data and Business Analytics: A Research Agenda For Realizing Business Value. Elsevier (2020).
30. Sun Z, Strang K, Li R. Big data with ten big characteristics. In: *Proceedings of the 2nd International Conference on Big Data Research*. Springer. (2018). (pp. 56–61).
31. Hasan MM, Popp J, Oláh J. Current landscape and influence of big data on finance. *J Big Data*. (2020) 7:1–17. doi: 10.1186/s40537-020-00291-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpubh.2022.905265

Integrated Blockchain-Deep Learning Approach for Analyzing the Electronic Health Records Recommender System

Eric Appiah Mantey[1]*, *Conghua Zhou*[1], *S. R. Srividhya*[2], *Sanjiv Kumar Jain*[3] and *B. Sundaravadivazhagan*[4]

[1] School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China, [2] Sathyabama Institute of Science and Technology, Chennai, India, [3] Department of Computer Science, Medi-Caps University, Indore, India, [4] Department of Information Technology, Faculty of Information Technology, University of Technology and Applied Sciences-Al, Mussanah, Oman

Blockchain is a recent revolutionary technology primarily associated with cryptocurrencies. It has many unique features including its acting as a decentralized, immutable, shared, and distributed ledger. Blockchain can store all types of data with better security. It avoids third-party intervention to ensure better security of the data. Deep learning is another booming field that is mostly used in computer applications. This work proposes an integrated environment of a blockchain-deep learning environment for analyzing the Electronic Health Records (EHR). The EHR is the medical documentation of a patient which can be shared among hospitals and other public health organizations. The proposed work enables a deep learning algorithm act as an agent to analyze the EHR data which is stored in the blockchain. This proposed integrated environment can alert the patients by means of a reminder for consultation, diet chart, etc. This work utilizes the deep learning approach to analyze the EHR, after which an alert will be sent to the patient's registered mobile number.

Keywords: electronic health records, blockchain, deep learning, integrated environment, hyperledger fabric

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Jude Osamor,
Glasgow Caledonian University,
United Kingdom
Praise Young,
Nnamdi Azikiwe University, Nigeria
Thippa Reddy Gadekallu,
VIT University, India

***Correspondence:**
Eric Appiah Mantey
ericphd@yahoo.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *26 March 2022*
Accepted: *04 April 2022*
Published: *06 May 2022*

Citation:
Mantey EA, Zhou C, Srividhya SR, Jain SK and Sundaravadivazhagan B (2022) Integrated Blockchain-Deep Learning Approach for Analyzing the Electronic Health Records Recommender System. Front. Public Health 10:905265. doi: 10.3389/fpubh.2022.905265

1. INTRODUCTION

In this modern period, many clinicians do not maintain the physical ledger of a patient's information, diagnostic details, and treatments. Instead, they keep health records electronically. These records hold the patient's complete information such as his history, allergies, treatment taken, etc. It can be viewed by clinicians, hospitals, and healthcare providers for any cause. These records require the utmost safety and security in order to prevent misuse by third parties or cyber attackers (1, 2). It contains confidential information about medication, symptoms, scan reports, and even biometrics. This work uses blockchain-based storage for Electronic Health Records (EHR) to provide privacy and security for the data. It is very crucial that EHR be maintained because it contains sensitive information (3, 4). There are some drawbacks in existing EHR maintenance which include:

1. Whether to trust an organization that stores EHR.
2. Hacking and medical data breaching.
3. Difficulty in retrieving the patient's history for treatment suggestion.
4. In case of emergencies, how to get the patient data, and if it can be trusted.

To overcome all these drawbacks, the work proposes a storage method using a blockchain-hyperledger fabric to store the patient data with enhanced security and privacy. The medical report of any patient cannot be accessed without their permission. Other users like doctors, healthcare providers should have a digital certificate from the government to join and access EHR in the blockchain network. Only doctors, healthcare providers, and clinicians would be able to perform CRUD operations in the block. The data from the block is retrieved using the hash value. Next, the retrieved patient's block is analyzed using Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM). Then the performance of RNN-LSTM is compared with RNN-GRU. Once the patient's block is analyzed, an alert will be sent to the patient's registered mobile number.

1.1. Motivation

The Electronic Health Records contain very sensitive information, and the chances of cyber attackers misusing the reports are very high in the centralized storage environment. Furthermore, the patients track their reports for regular consultation, medication, diagnosis appointments, etc. The motivation of this proposed work is to analyze the patient's EHR witha deep learning mechanism stored in the Blockchain and to create an alert system for the patient.

1.2. Contribution

To incorporate the reliable Hyperledger fabric blockchain to store the medical records of the patient. It also analyzes the blocks(the patient's EHR) using the RNN-LSTM and RNN-GRU mechanism. An alert system is created for the patients welfare, for activities like an alert for the next consultation, medication, and for diagnosis. The behavior and efficiency of the RNN algorithm is analyzed. Asa result, a reliable Hyperledger fabric is utilized for storing the EHR. Those medical records are analyzed by the deep learning Recurrent Neural Network algorithm, and it creates an alert system to monitor the patients remotely.

1.3. Organization of the Paper

The rest of the paper is organized as follows: Section 2 describes related works. Section 3 explains materials and methods. Section 4 discusses the results. Finally, Section 5 concludes the work.

2. RELATED WORK

Blockchain is a booming technology which has been incorporated in many areas like industry and academic scenarios (3, 5). The diet recommendation system was created for patients affected by heart disease and to avoid heart disease (6), while also checking for the family history. Kutia et al. (7), Plastiras and O'Sullivan (8), and Shabbir et al. (9) studied the impact of different parameters that influence the adapting to and use of the online health facilities in Ukraine and China, such asallergies, food preferences, age, blood pressure. It utilizes machine learning algorithms for the recommendation system. The authors (8) proposed a tailored recommendationfor patients based on different parameters, with reference to their health records using Artificial Intelligence.

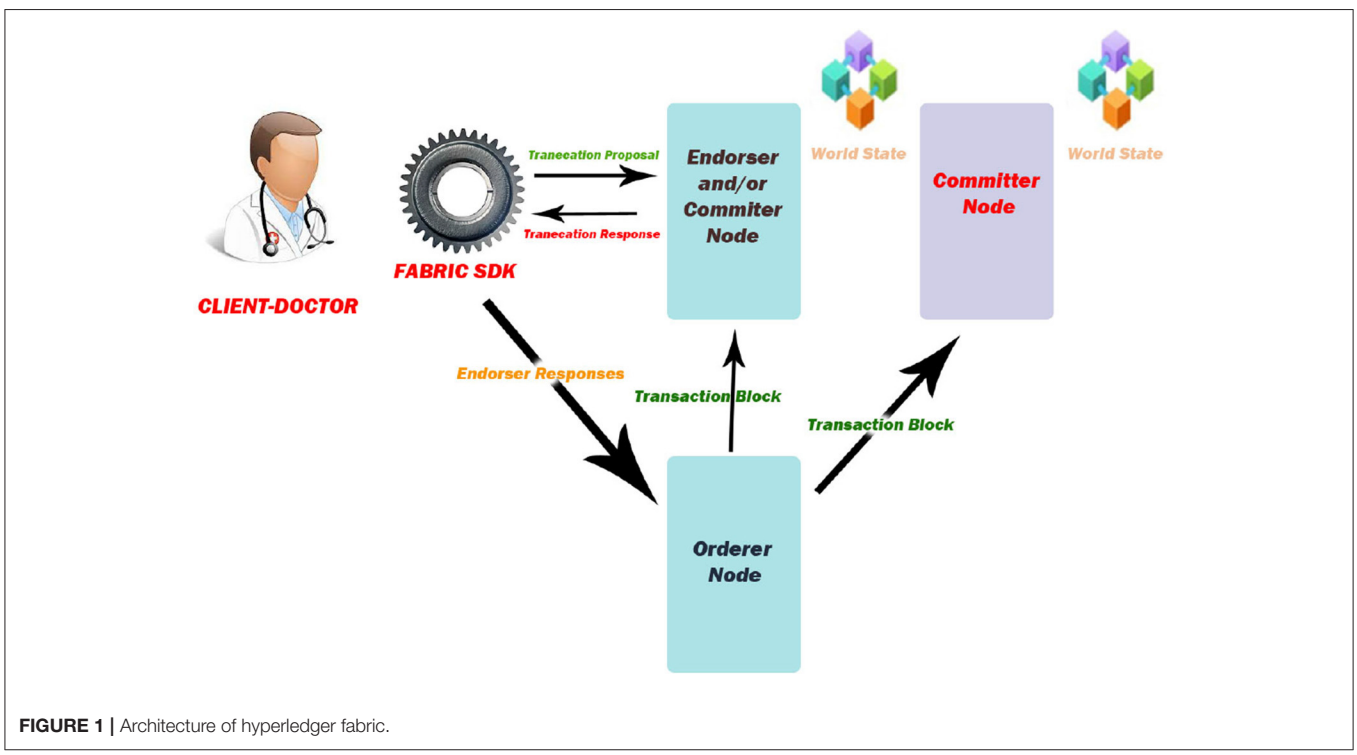


FIGURE 1 | Architecture of hyperledger fabric.

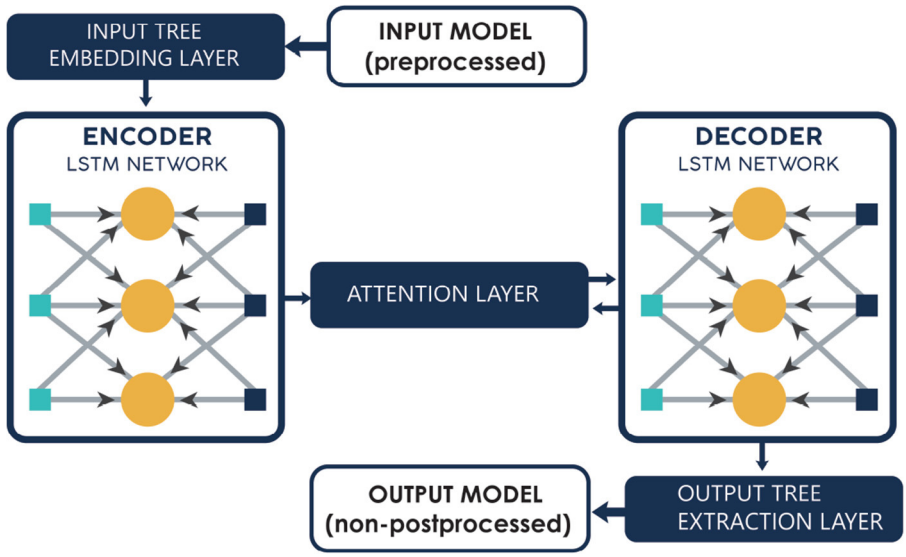


FIGURE 2 | Architecture of the LSTM RNN.

FIGURE 3 | Model of the system.

TABLE 1 | Reference summary.

S.no	Health information system	Challenges	Area to be concentrated
1	Electronic health records [21]	Interoperability, access control, data integrity	Shared decision making
2	Electronic health records [22]	Interoperability	Health data recording, storing, and sharing Access control
3	Electronic health records [23]	Interoperability	Sharing of healthcare information for clinical and research purposes Access control
4	Electronic health records [24]	Data integrity, access control	Sharing healthcare data between health institutions
5	Electronic health records [25]	Data integrity	Sharing healthcare data between health institutions
6	Electronic health records [26]	Access control, interoperability, data integrity	Sharing healthcare data for clinical and research purposes
7	Electronic health records [27]	Access control, interoperability	Sharing healthcare data for clinical and research purposes
8	Electronic health records [28]	Access control, interoperability and administrative	Sharing healthcare (health record) information for clinical, research[economic] purposes.
9	Electronic health records [29]	Access control, data integrity, interoperability	Patients collection, archiving and sharing of healthcare data for clinical purposes
10	Electronic health records [30]	Patient data management and storage	Environment Access control, data integrity, data provenance

The study (10) proposes a literature survey of machine learning algorithms that are used in a recommendation system. Agapito et al. (11) proposed a diet organizer system, which can be utilized by both healthy people and patients with chronic diseases. The authors in Mani et al. (12) and Singhal et al. (13) utilized the blockchain, Hyperledger fabric storage using Inter Planetary File Storage (IPFS) protocol to save data in the blockchain. Wang et al. (14) showed how bigdata analytics can be utilized for

the healthcare industry, with the pros and cons of the model. The use of predictive data mining algorithms to recommend healthy diets was proposed by Jaiswal (15). A data mining model was developed to provide users with information about healthy food habits and eating patterns, such as the number of calories burned, the amount of macronutrients consumed, and so on. Patients' dietary preferences are predicted by the patient diet recommendation system by taking into account their eating habits and body measurements. In spite of its effectiveness in predicting healthy diets for patients and nutritionists/doctors, the study is limited by its lack of a flexible model and minimally tailored solutions to fit the patient's needs. Argaw et al. (16) briefly explains that clinical documents and hospitals are very prone to hackers and cyber attackers. There are scalability issues as the data is stored in an on-chain database, which makes the system unstable. However, the author in Egala et al. (17) proposed a ring structure-based access control that ensures privacy. Alsufyani et al. (18) proposed a system to manage data intelligently in a cyberspace environment. Decentralized storage and access to records, as described in Singh et al. (19), are efficient ways to utilize the network's power and resources. In Peng et al. (20), the author employed high-end privacy-enhancing technologies, such as homomorphic encryption, that prevent vulnerabilities due to the processing of data while it is encrypted. To ensure nodes did not engage in malicious behavior, the Agrawal and Jain (21) used zero-knowledge proof and proof authority consensus for mutual authentication as a privacy enhancement technique. There is no doubt that blockchain technology provides a cryptographic solution to the security issue, but there are challenges such as privacy, scalability, and interoperability. Large medical data centers have experienced record breach episodes in the past two decades, which have left medical companies with additional challenges (21). When blockchain technology was newly invented, MedRec (22) became the first recommendation system for an electronic patient record management system that would be made possible with blockchain technology. Ether blockchains are capable of backing up detailed accessibility data. Healthcare providers maintain third-party databases, not the blockchain. This means that the records can still be hacked or misused. A healthcare management system as in Ivan (23) encrypts patient keys as data is recorded on the blockchain. Researchers and hospitals decrypt the data with a patient's public key after receiving consent from the patient. In contrast, we empower our patients to control their own data, so they are the only ones who have access to them.

3. MATERIALS AND METHODS

3.1. Hyperledger Fabric

Hyperledger fabric is an open-source platform which builds the distributed ledger, with a plug-and-play architecture (modular) that facilitates a high level of security, privacy, and confidentiality of the data. The architecture of Hyperledger fabric is shown in **Figure 1**. The client capitulates the transaction pool through the fabric SDK to the endorser. Endorsing peers verify and execute the transaction, and generate the read and write sets. The response is then sent back to the client. All replies from peers are collected by the client, and they are sent to the "orderer." In this case, all transactions are ordered by the orderer in ascending order and then formed into a block. Each committer validates this block and adds a new block to their own copy of the ledger as a result. The fabric consists of four components, namely:

1. Membership Service Provider (MSP)
2. Client
3. Orderer
4. Peer

Membership Service Provider is the main component of Fabric which defines a rule to allow members in after authentication and verification. It manages the User ID and the authenticity of the client. The client is the person who initiates the transaction proposal. All transactions have to be written in the shared ledger in blockchain in a consistent order. In order for updates to the world state to be valid, the order in which they are made must be determined.

TABLE 2 | List of abbreviations.

HR	Health Record
P1	Patient
D1	Doctor
U1	User
PPK	Patients Private key
PUK	Patients Public Key
Sk	Session Key
DPK	Doctors Private key
PV	Patient View of data

3.2. RNN-Long Short-Term Memory

A Recurrent Neural Network is a special kind of deep learning where the output of the previous step will be the input of the subsequent step. LSTM is a special type of Recurrent Neural Network which enables the system to learn the long-term dependencies of data. This type of learning is achieved by the repeated module of the LSTM, which has a combination of 4 different layers connected to each other. **Figure 2** explains the method, and the three-layeredarchitecture where the embedding layer gives the pre-processed data to the Encoder-LSTM network. It then passes to the attention layer that acts as the intermediate layer, where it facilitates the decoder to pay more attention to the specific parts of the fixed-size vectors.

3.2.1. RNN-GRU

The Gated Recurrent Unit (GRU) is another type of Recurrent Neural Network. It is very similar to the Long-Short-Term Memory RNN.

3.2.2. System Model

This work utilizes the blockchain to store all the EHR of the patients, which cannot be accessed by third parties or cyber attackers, because of its high security and privacy of medical data. Hyperledger Fabric can be used to save these types of

medical data. It uses Inter Planetary File System (IPFS) which provides a solution for the file storage problems. IPFS can store and retrieve large files efficiently. Health records are encrypted using symmetric key cryptography to ensure privacy. To protect the record, it is stored in an encrypted format on an IPFS server under the appropriate supervisory authority. To access the patient's record, an entity must first be given the authorization to do so.

1. A private key is used to decrypt the record.
2. RSA key pair public key and symmetric key are used to encrypt the key.

Access to a health record may be removed if:

1. The symmetric key is decrypted by the private key associated with the EHR owner.
2. The symmetric key is used to decrypt the EHR.
3. A new symmetric key is used to re-encrypt the record.
4. Encryption of the symmetric key is completed using the public keys of all authorized users.

Figure 3 describes the model of the system it takes the Input (EHR) from clinicians, doctors, and healthcare providers which is stored in the Hyperledger Fabric using the protocol Inter Planetary File System (IPFS) using **Algorithm 1**, where from the data is retrieved by a deep learning agent for analyzing the EHR, after analyzing the data in EHR the RNN-LSTM will send the alert for several activities like doctor's consultation, medication, and diagnosis schedule to the patient's registered android device, this was explained in the **Algorithm 2**.

This proposed work uses nine different features from EHR to provide an effective alert system. These features are listed in **Table 1** above. Considering all these features, we can design an efficient alert system, through whichregular monitoring of thepatient can be done.

Algorithm 1: Creating and updating health records in Hyperledger blockchain.

Storing and updating data ()
Input: (HR) from various sources
output: 0 or 1

1. Procedure Store_EHR
2. For each
3. U1 access P1
4. If (permission== GRANT && role== DOCTOR) then
5. Create PV of HR in IPFS
6. PV Decryption (Encrypt (HR))
7. Create Sk
8. send Encrypted (PPK (Sk), DPK(Sk), PV(Sk)) to P1, D1, and PV
9. create_Update ()
10. HR[(Decryption PPK (Encrypted PUK (HR) + Encryption (U)]
11. Pan Commit (IPFS (HR))
12. IPFS HR_hash
13. HR_hash HyperlegerFabric Blocks
14. Return 1
15. Else
16. Permission=Deny
17. Return 0
18. End if
19. End For ()
20. End procedure Store_EHR

3.3. Data Preprocessing

Data pre-processing is an essential task to be carried out. It helps to achieve the maximized efficiency of the model. At this stage, the dataset is pre-processed by four important steps which include:

1. Data cleaning
2. Data integration
3. Data transformation
4. Data reduction

These four steps have to be followed sequentially to get the required pre-processed data to train the model. At first, the dataset is cleansed by assigning missing values, null values, eliminating the nosy data, resolving the inconsistency,and removing outliers. The cleansed data is then used for data integration, which refers to the merging of data from multiple sources into a single larger data store, such as a data warehouse. Inthe data transformation step, wetransform the value, structure, or format of quality data into alternative forms, according to some strategies such as generalization, normalization, attribute selection and aggregation. Finally, the data reduction is carried out to find the appropriate dataset model proposed to achieve the maximum efficiency.

3.4. Evaluation Metrics

Many evaluation metrics are used to evaluate the performance of our model

$$Accuracy = (AZ + AN)/(AZ + AN + CZ + CN) \quad (1)$$

$$Precision = AZ/(AZ + CZ) \quad (2)$$

The objective of precision is to examine the True Positive (AZ) units in connection with False Positive (CZ) units.

The objective of recall is to examine True Positive (AZ) units in connection to False Negative (CN) units that are not classified. The arithmetic arrangement of recall is stated in Equation below:

$$Recall = AZ/(AZ + CN) \quad (3)$$

Sometimes, the assessment of performance may not be very efficient with recall and accuracy. For example, if a mining algorithm has high precision but low recall, then another algorithm is needed. Then comes the question of which algorithm is more effective. This challenge is solved utilizing F1-measure

Algorithm 2: Creating an alert for health records in Hyperledger blockchain Alert ().

Input: HR from BLOCKCHAIN
output: Recommendation in Mobile
Procedure Recommendation
Initialize V,U;
Preprocess content of items;
For I to N do
    Foreach (X_{sample},V_{sample}) ϵ samples do
        trainAutoencoder (X_{sample},V_{sample});
    end
    theta $\leftarrow$ Autoencoder(X);
    $U_tU \leftarrow U^TU$;
    For each I ϵ items do
        $U_tU \leftarrow U_tU + \alpha U_{OB} *^T U_{OB} + \lambda_V I$;
        $U_{tr} \leftarrow (1+\alpha)\ U_{OB}r_{OB} + \lambda_v$theta
        $V_i \leftarrow$ Solve (U_tU,Utr$_{ob}$)
    End
    VtV $\leftarrow V^TV$;
    Foreach u ϵ users do
        VtV $\leftarrow$ VtV+$\alpha\ V_{OB} *^T V_{OB} + \lambda_U$I;
        $V_{tr} \leftarrow (1+\alpha)\ V_{OB}r_{OB}$
        $U_U \leftarrow$ Solve (U_tU,Utr$_{ob}$)
    End
End

that gives a mean recall and precision. F1-measure can be calculated as follows:

$$F1score = 2 * (Precision * Recall)/(Presiion + Recall) \quad (4)$$

TABLE 3 | Features used in the model.

S. No	Features	Type of feature values
1	Patient_ID	Numeric
2	Patient_age	Numeric
3	Patient_Gender	Categorical
4	Patient_Weight	Numeric
5	Patient disease	Categorical
6	Medication chart	Text
7	Next Appointment date	Date
8	Next diagnosis date	Date
9	Diet specification	Text

TABLE 4 | Report of precision, recall, and F1 scores.

Parameters	Label	Integrated hyperledger Fabric-LSTM	Integrated hyperledger Fabric-GRU
Precision	Allowed	0.9876	0.9654
	Not allowed	0.8943	0.9923
Recall	Allowed	0.9946	0.9986
	Not allowed	0.7231	0.3743
F1 Scores	Allowed	0.9921	0.9723
	Not allowed	0.8276	0.5523

3.5. Experiments and Results

In the proposed model, the personal alert system is designed for patients registered mobile numbers. The alert is created for certain activities like the next consultation date, medication, diet specification, and diagnosis date. Initially, the patient's EHR is stored in the Blockchain using IPFS protocol. The data saved in the blockchain will be more secure and distributed. The

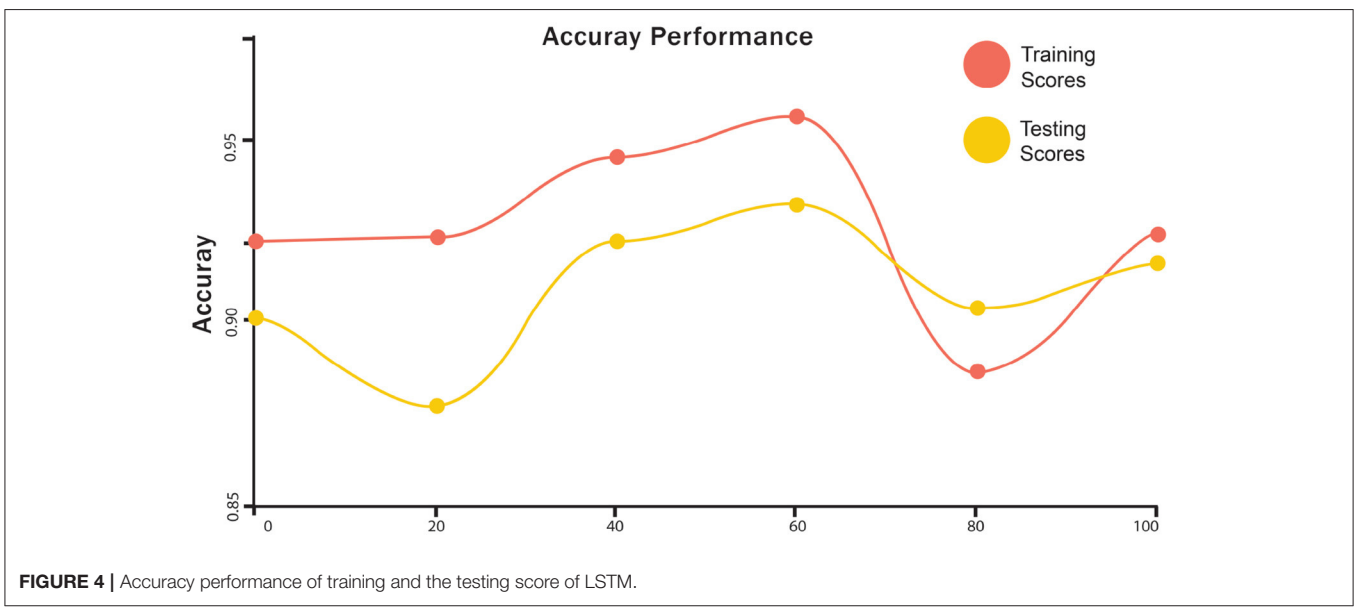


FIGURE 4 | Accuracy performance of training and the testing score of LSTM.

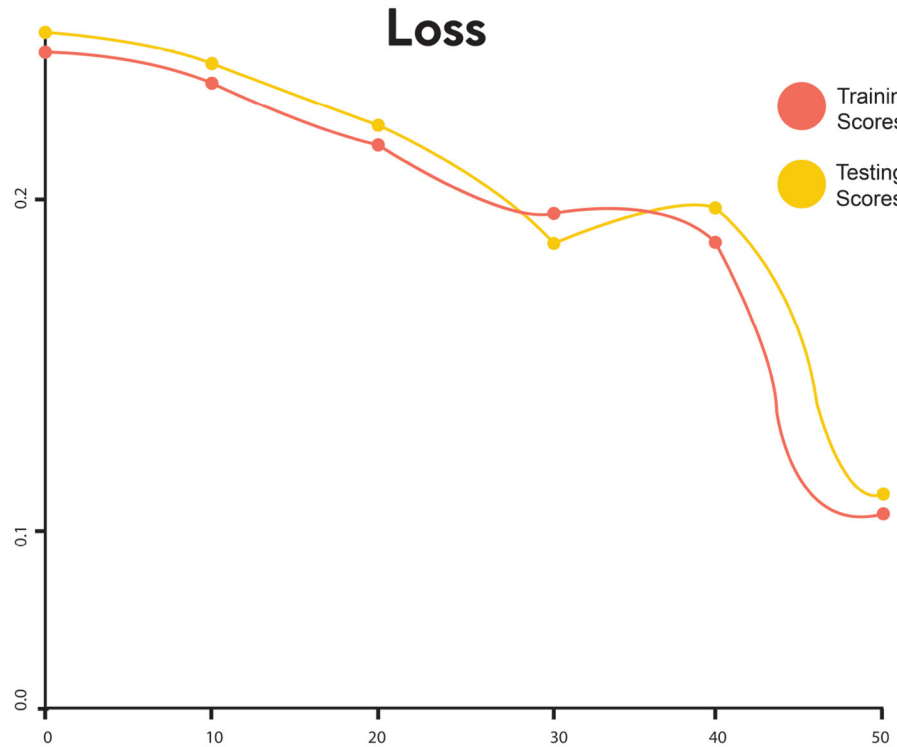


FIGURE 5 | Loss performance of training and the testing score of LSTM.

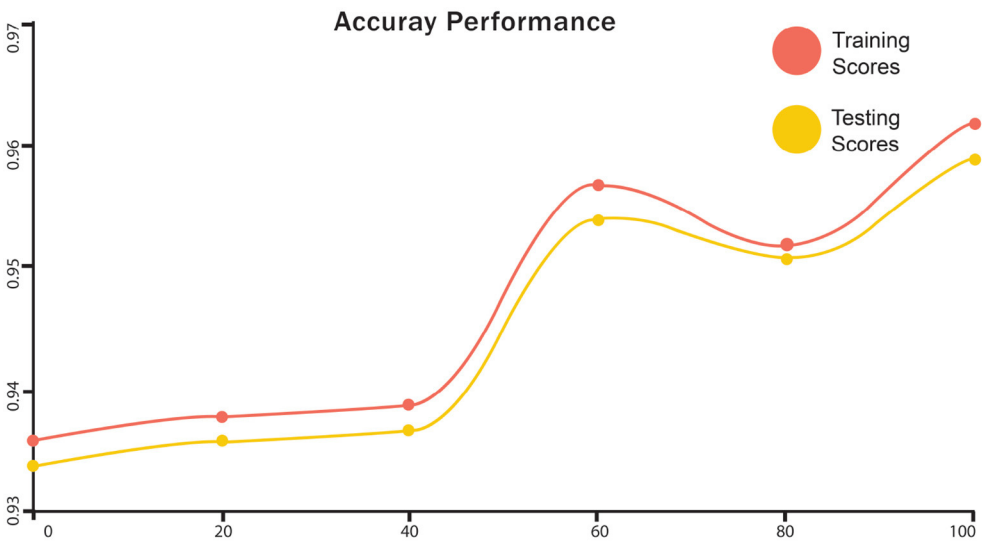


FIGURE 6 | Accuracy performance of training and testing score of GRU.

data that is saved in the blockchain is retrieved by the deep learning mechanism, Long Short-Term Memory for analyzing the data. After analyzing the patient EHR, a tailored alert system is triggered to the registered mobile number.

The proposed work uses nine features from the dataset, the features are ID, age, gender, disease, weight, consultation date, medication, diagnosis, and diet specification. The taxonomical analysis is done with several featuresto give accurate results.

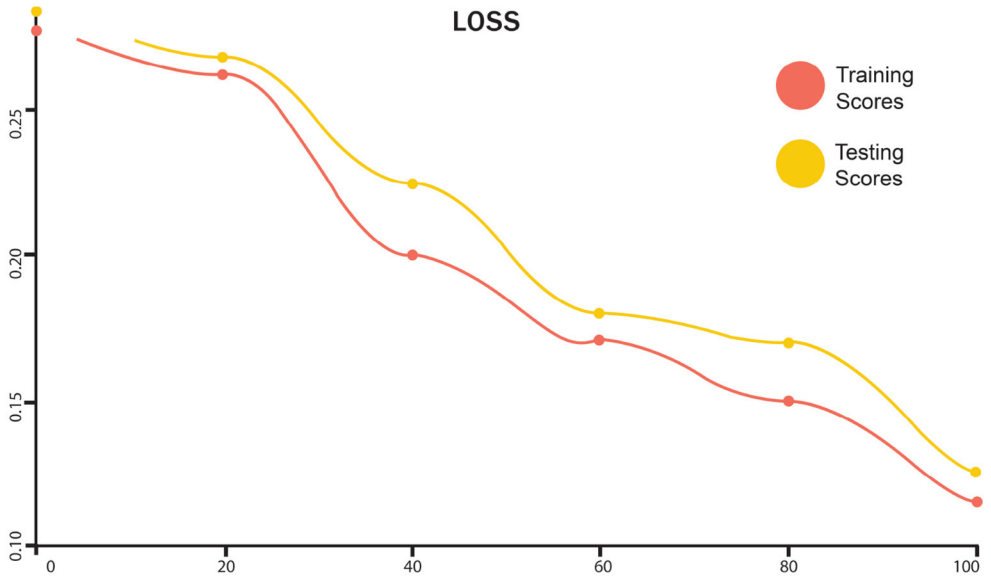


FIGURE 7 | Loss performance of training and testing score of GRU.

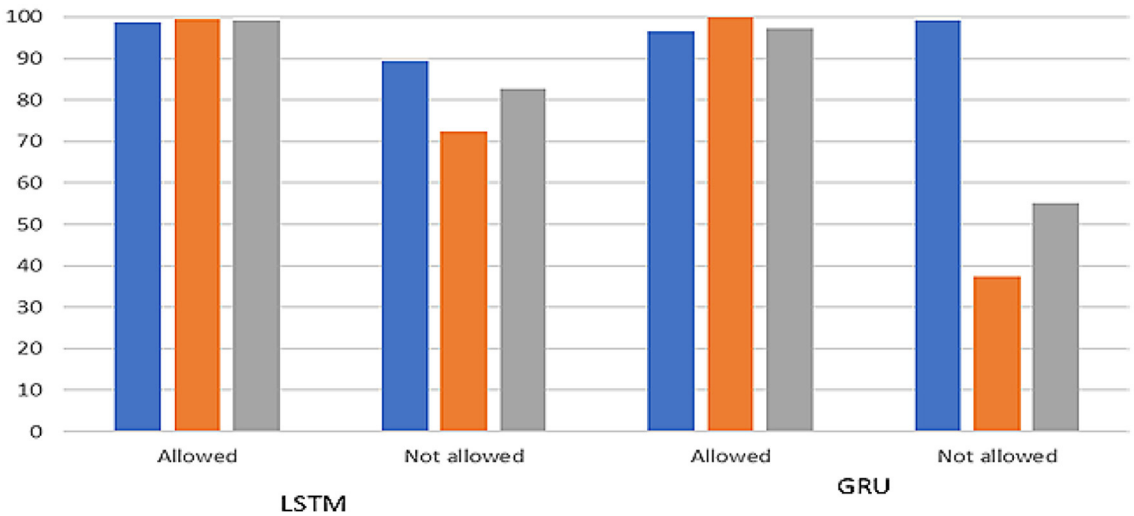


FIGURE 8 | Classification report of LSTM and GRU.

The analysis of the EHR is done using the deep learning techniques Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) which are Recurrent Neural Network techniques. **Table 2** shows the list of abbreviation and **Table 3** shows the features used in the model. Therefore, the comparison between the two integrated blockchain and Recurrent Neural Networks in analysing the EHR is given in **Table 4**, in terms of precision, recall, and F1 score.

The training and testing scores for LSTM are shown in **Figure 4**. In the graph, the red line represents the training curve and the yellow line represents the testing curve. Training curves start from 93.6% and after 50 epochs, they reach 95%. For the testing curve, it starts at 94% and goes to 97.4%, then goes down to 96%. Training and testing losses for LSTM are shown in **Figure 5**. The training loss is represented in red, while the testing loss is displayed in yellow. The training loss begins at 0.27 and reduces to 0.128. The testing loss begins at 0.23 and decreases to 0.075. **Figure 6** shows the scores obtained during the training and testing for GRU, while **Figure 7** shows the losses obtained during the training and testing for GRU. The red curve in **Figure 5** represents the training curve. It starts at 93% and goes up to 97% testing and training scores for GRU. Similarly, the yellow curve representing the LSTM's testing score starts at 94% and goes up to 97%. **Figure 7** shows the training loss starts at 0.3 and decreases

after 50 iterations. It reaches 0.1 after that. Likewise, the testing loss starts at 0.2 and decreases until it reaches 0.05.

Table 2 shows that the LSTM model outperforms another modelin terms of recall, precision, and F1 score. The Allowed LSTM had 98% precision, 99% recall, and F1 measure scores, respectively. The Not Allowed class had 87% precision,77% recall, and 81% F1 measure score, respectively. The other model Integrated Blockchain-GRU also performs well but not as well as LSTM. Finally, the result of the model is shown for both Allowed and Not Allowed classes in **Figure 8**.

4. CONCLUSION

The study proposes an integrated blockchain-RNN algorithm for storing the EHR in the Hyperledger Fabric using the protocol Inter Planetary File System. The stored EHR is analyzed with the deep learning mechanism, Recurrent Neural Network algorithms, namely Long-short-Term Memory and Gated Recurrent Units. The integrated model gives an alert to the registered mobile number of patients regarding the consultation reminder, diagnosis alert, medication, diet chart specification. It uses 9 features, i.e., age, sex, weight, disease, medication chart, appointment date, diagnosis date, and diet specification. This proposed work focuses on an integrated model to automate the alert system for various activities of the patient. Finally, experimental results show that the LSTM outperforms the other models in terms of precision, recall, and F1 score. This work is practically possible but the maintenance cost is more when compared to the traditional model. In the Future, the alert system can be improved by collaborating with the calendar application in android mobile with fitness applications and to provide a solution for a cost-efficient model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This article was supported by the key research and development plan (social development) projects BE2016630 and BE2017628 of Jiangsu province, the scientific research project Z201603 of Wuxi health and family planning commission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.905265/full#supplementary-material

REFERENCES

1. Mantey EA, Zhou C, Anajemba JH, Okpalaoguchi IM, Chiadika ODM. Blockchain-secured recommender system for special need patients using deep learning. *Front Public Health*. (2021) 9:737269. doi: 10.3389/fpubh.2021.737269
2. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst*. (2021) 28:1–15. doi: 10.1007/s00530-021-00774-w
3. Bhattacharya P, Tanwar S, Bodkhe U, Tyagi S, Kumar N. Bindaas: blockchain-based deep-learning as-a-service in healthcare 4.0 applications. *IEEE Trans Netw Sci Eng*. (2019) 8:1242–55. doi: 10.1109/TNSE.2019.2961932
4. Iwendi C, Anajemba JH, Biamba C, Ngabo D. Security of things intrusion detection system for smart healthcare. *Electronics*. (2021) 10:1375. doi: 10.3390/electronics10121375
5. Ngabo D, Wang D, Iwendi C, Anajemba JH, Ajao LA, Biamba C. Blockchain-based security mechanism for the medical data at fog computing architecture of internet of things. *Electronics*. (2021) 10:2110. doi: 10.3390/electronics10172110
6. Kim JH, Lee JH, Park JS, Lee YH, Rim KW. Design of diet recommendation system for healthcare service based on user information. In: *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*. Seoul: IEEE (2009). p. 516–8.
7. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-Technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access*. (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
8. Plastiras P, O'Sullivan D. Exchanging personal health data with electronic health records: a standardized information model for patient generated health data and observations of daily living. *Int J Med Inform*. (2018) 120:116–25. doi: 10.1016/j.ijmedinf.2018.10.006
9. Shabbir A, Shabbir M, Javed AR, Rizwan M, Iwendi C, Chakraborty C. Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals. *J Exp Theor Artif Intell*. (2022) 1–28. doi: 10.1080/0952813X.2021.1960634
10. Portugal I, Alencar P, Cowan D. The use of machine learning algorithms in recommender systems: a systematic review. *Expert Syst Appl*. (2018) 97:205–27. doi: 10.1016/j.eswa.2017.12.020
11. Agapito G, Calabrese B, Guzzi PH, Cannataro M, Simeoni M, Caré I, et al. DIETOS: a recommender system for adaptive diet monitoring and personalized food suggestion. In: *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. New York, NY: IEEE (2016). p. 1–8.
12. Mani V, Manickam P, Alotaibi Y, Alghamdi S, Khalaf OI. Hyperledger healthchain: patient-centric IPFS-based storage of health records. *Electronics*. (2021) 10:3003. doi: 10.3390/electronics10233003
13. Singhal V, Jain S, Anand D, Singh A, Verma S, Rodrigues JJ, et al. Artificial intelligence enabled road vehicle-train collision risk assessment framework for unmanned railway level crossings. *IEEE Access*. (2020) 8:113790–113806. doi: 10.1109/ACCESS.2020.3002416
14. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol Forecast Soc Change*. (2018) 126:3–13. doi: 10.1016/j.techfore.2015.12.019
15. Jaiswal V. A new approach for recommending healthy diet using predictive data mining algorithm. *Int J Res Anal Rev*. (2019) 6:58–65.
16. Argaw ST, Bempong NE, Eshaya-Chauvin B, Flahault A. The state of research on cyberattacks against hospitals and available best practice recommendations: a scoping review. *BMC Med Inform Decis Mak*. (2019) 19:1–11. doi: 10.1186/s12911-018-0724-5

17. Egala BS, Pradhan AK, Badarla V, Mohanty SP. Fortified-chain: a blockchain-based framework for security and privacy-assured internet of medical things with effective access control. *IEEE Internet Things J*. (2021) 8:11717–31. doi: 10.1109/JIOT.2021.3058946
18. Alsufyani A, Alotaibi Y, Almagrabi AO, Alghamdi SA, Alsufyani N. Optimized intelligent data management framework for a cyber-physical system for computational applications. *Complex Intell Syst*. (2021) 1–13. doi: 10.1007/s40747-021-00511-w
19. Singh P, Masud M, Hossain MS, Kaur A. Blockchain and homomorphic encryption-based privacy-preserving data aggregation model in smart grid. *Comput Electr Eng*. (2021) 93:107209. doi: 10.1016/j.compeleceng.2021.107209
20. Peng C, He D, Chen J, Kumar N, Khan MK. EPRT: an efficient privacy-preserving medical service recommendation and trust discovery scheme for eHealth system. *ACM Trans Internet Technol*. (2021) 21:1–24. doi: 10.1145/3397678
21. Agrawal S, Jain SK. Medical text and image processing: applications, issues and challenges. In: *Machine Learning with Health Care Perspective*. Cham: Springer (2020). p. 237–62.
22. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: using blockchain for medical data access and permission management. In: *2016 2nd International Conference on Open and Big Data (OBD)*. Vienna: IEEE (2016). p. 25–30. doi: 10.1109/OBD.2016.11
23. Ivan D. Moving toward a blockchain-based method for the secure storage of patient records. In: *ONC/NIST Use of Blockchain for Healthcare and Research Workshop*. Gaithersburg, MD: United States: ONC/NIST (2016). p. 1–11.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mantey, Zhou, Srividhya, Jain and Sundaravadivazhagan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpubh.2022.897535

Effect of Network-Based Positive Psychological Nursing Model Combined With Elemene Injection on Negative Emotions, Immune Function and Quality of Life in Lung Cancer Patients Undergoing Chemotherapy in the Era of Big Data

Shilu Yang[1†], *Lijuan Zheng*[2†], *Yan Sun*[3] *and Zhuoyun Li*[4*]

[1] Intensive Care Unit, Wuhan Jinyintan Hospital, Wuhan, China, [2] Department of Drug Resistant Tuberculosis, Wuhan Jinyintan Hospital, Wuhan, China, [3] Jimo Nanquan Health Center, Qingdao, China, [4] Department of Pharmacy, Qingdao Municipal Hospital, Qingdao, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Rmahammad Shafi,
Mizan Tepi University, Ethiopia
Mohammad Monirujjaman Khan,
North South University, Bangladesh

***Correspondence:**
Zhuoyun Li
lizhuoyun@qdslyy.com.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 16 March 2022
Accepted: 04 April 2022
Published: 06 May 2022

Citation:
Yang S, Zheng L, Sun Y and Li Z (2022) Effect of Network-Based Positive Psychological Nursing Model Combined With Elemene Injection on Negative Emotions, Immune Function and Quality of Life in Lung Cancer Patients Undergoing Chemotherapy in the Era of Big Data. Front. Public Health 10:897535. doi: 10.3389/fpubh.2022.897535

Background: With the development of big data, big data interpenetrate in every person's life. Health care is no exception to this trend, especially in regard to nursing analytics. The data that is analyzed and leveraged in this field is gathered from a variety of sources, including electronic health records (EHRs), medical histories, provider notes and mobile applications, creating an accumulation of personalized health information around each individual.

Objective: To explore the effect of the network-based positive psychological nursing model combined with elemene injection on negative emotions, immune function and quality of life (QOL) in patients with lung cancer (LC) undergoing chemotherapy.

Methods: The clinical data of 90 LC patients who underwent chemotherapy in our hospital from December 2020 to December 2021 were retrospectively analyzed, and the patients were equally split into experimental group (EG) and control group (CG) according to the order of enrollment. The patients in CG received routine nursing intervention during chemotherapy, while those in EG received the network-based positive psychological nursing model combined with elemene injection to compare negative emotions, immune function and quality of life (QOL) between the two groups.

Results: Compared with CG, EG had notably higher immune function indexes ($P < 0.001$), lower serum VEGF and MMP-2 levels ($P < 0.001$), lower HAMA and HAMD scores ($P < 0.05$), and higher scores of PTGI, physiological function, psychological function, social function and physical function ($P < 0.001$).

Conclusion: The network-based positive psychological nursing model combined with elemene injection is a reliable method to enhance the immune function and QOL of LC patients undergoing chemotherapy and alleviate their negative emotions, which has a high clinical application value.

Keywords: network, positive psychology, elemene injection, lung cancer chemotherapy, negative emotions, immune function, quality of life (QOL)

INTRODUCTION

Lung cancer (LC), mainly occurring in bronchial epithelial cells, is a common clinical malignant tumor with a high incidence (1) and seriously threatens human health. Thoracoscopic radical resection of lung cancer is a common surgical approach to treat LC (2), and postoperative chemotherapy is an indispensable adjuvant therapy. However, chemotherapy can trigger serious adverse reactions in patients, such as gastrointestinal reactions, vomiting and diarrhea (3, 4). In addition, patients during the period of chemotherapy often experience negative emotions due to worries about postoperative recovery and fear of tumors. In addition, the prevalence of the concept of cancer fear and the harm of LC itself expose patients to great psychological pressure, such as anxiety and tension. Besides, studies have found that immune function plays an important role in the disease development of malignant tumor patients, and cancer cells seriously affect the immune function of patients after the occurrence of tumors. At the same time, the toxic reactions of chemotherapy drugs also damage the normal physiological function of patients and result in a series of severe stress reactions, further damaging the immune system of patients, resulting in the decline of immune function and seriously affecting the quality of life (QOL). Therefore, how to alleviate the negative emotions of LC patients undergoing chemotherapy, enhance the immune function and QOL, and prevent disease recurrence is an urgent problem to be solved in clinic (5, 6). Routine nursing measures are carried out from health education, cognitive management and emotional support, which can hardly stimulate the potential of patients and leaves an unsatisfactory effect. Based on the theory of positive psychology, positive psychological intervention taps individual positive factors and stimulates the intrinsic potential of patients through their positive emotional experience and effective social support (7, 8). The development of Internet technology effectively solves the current lack of nursing resources in China. A report (9) shows that the network-based positive psychological nursing model enables patients to understand their diseases and maximizes the effect of psychological intervention through the instant push function of mobile phone APPs. A current study (10) has found that elemene injection can reduce the mitotic ability of tumor cells and induce their apoptosis, thus inhibiting their growth and proliferation, changing their immunogenicity, and strengthening the immune response of the body. Chemotherapy causes great physical and psychological pain to LC patients. In order to alleviate the pain of such patients, this study can play a guiding role in the clinical treatment of lung cancer, effectively improve the negative emotions of patients undergoing chemotherapy, reduce the economic burden of their families, and provide theoretical support for further research in the future. At present, few studies have confirmed the intervention effect of the network-based positive psychological nursing model combined with elemene injection on LC patients undergoing chemotherapy. In order to fill this gap and provide evidence-based evidence for more patients, this study is reported as follows.

MATERIALS AND METHODS

General Information

The study subjects were 90 LC patients who received chemotherapy in our hospital from December 2020 to December 2021, and were equally split into experimental group (EG) and control group (CG) according to the order of enrollment. This study was approved by the ethics committee of *Jimo Nanquan Health Center*, following the Declaration of Helsinki (2013) (11).

Inclusion and Exclusion Criteria

Inclusion Criteria

(1) Patents meeting the diagnostic criteria of LC and receiving chemotherapy; (2) patents with no history of epilepsy or cognitive impairment; and (3) patents with the stable condition and the approval of the attending doctor.

Exclusion Criteria

(1) Patents with a deteriorating condition and unstable vital signs; (2) patents with inability to understand the purpose of the researchers and the content of the questionnaire, and with severe cognitive impairment; and (3) patents who were allergic to the research drugs.

Methods

Patients in CG received routine nursing intervention during chemotherapy through health education and professional guidance, including the pathogenesis, clinical manifestations and stages of disease development, precautions during chemotherapy, side effects of chemotherapy drugs, and dietary guidance (12, 13).

Patients in EG received the network-based positive psychological nursing model combined with elemene injection. The network-based positive psychological nursing model was carried out as below. (1) A network-based positive psychological nursing team was established, with 1 attending physician, 1 head nurse, 3 whole-course nurses, and 2 assistant nurses, aiming to organize the training on related knowledge of LC,

chemotherapy and positive psychology, and carry out the assessment. The nurses could take the post only when they passed the assessment. (2) Register on the mobile APP. The LC health management platform was established, including modules such as chemotherapy period, psychological communication and patient assistance. The patients were guided to enter the platform and each module. The head nurse was the platform administrator and the whole-course nurses were responsible for the specific operations. (3) Online knowledge learning. According to the relevant content of positive psychology, combined with the psychological problems of LC patients undergoing chemotherapy, the head nurse organized and compiled the network courses of positive psychological intervention. The whole-course nurses pushed the contents on the WeChat public account weekly to guide patients to learn and use positive psychology to face the discomfort after chemotherapy. (4) Sharing of positive emotions on the platform. The patients were encouraged to write gratitude logs and upload them to the platform to cultivate their positive attitude. (5) Online mindfulness decompression training. The whole-course nurses led the patients to carry out online mindfulness decompression training, and guided them to exchange their feelings and ideas and express their psychological pain. Elemene injection (manufacturer: CSPC Pharmaceutical Group Co., Ltd.; NMPA approval No.: H20110114; specification: 10 ml: 0.2 g × 5 vials/box) was injected into patients, 400 mg/time, 1 time/d. With 21 days as a course of treatment, the patients received the continuous treatment of three courses. The dosage was appropriately adjusted according to the actual situation of patients during the treatment process.

Observation Indexes

The fasting elbow venous blood of both groups after treatment was collected and centrifuged to prepare serum samples. The serum matrix metalloproteinases (MMP-2) and vascular endothelial growth factor (VEGF) were measured by ELISA. A protein immunoassay analyzer (model: CS-AutoBlot 48; manufacturer: Hunan Zhongrui Huxin Medical Technology Co., Ltd.) was used to detect the levels of immunoglobulin (Ig) G, IgA and IgM.

Hamilton Anxiety Scale (HAMA) (14) and Hamilton Depression Scale (HAMD) (15) were used to evaluate the negative emotions of both groups. The total score was 56 in HAMA and 68 in HAMD. Higher scores suggested worse emotions of patients.

Post-traumatic Growth Level

The Chinese version of Post-traumatic Growth Inventory (PTGI) (16) was applied to evaluate the post-traumatic growth level of patients, including life perception (4 items), personal strength (4 items), new possibility (3 items), relationship with others (6 items) and self-transformation (3 items). The Likert 6-level scoring method was adopted with 0–5 points from "no" to "very much", and a total score of 100 points. A higher score indicated a higher post-traumatic growth level.

Quality of Life (QOL)

The Quality of Life (QOL) scale was used to evaluate the patients from psychological function, social function, physical function and physiological function, with each scoring 100 points. A higher score denoted better QOL.

Statistical Methods

The data were processed by the professional statistical software SPSS26.0 and graphed by GraphPad Prism 7 (GraphPad Software, San Diego, USA). Enumeration data were tested by X^2 and expressed as [n (%)], while measurement data were tested by t-test and expressed as Mean ± SD. When $P < 0.05$, the differences were statistically significant.

RESULTS

Clinical Data

No obvious differences in sex ratio, times of chemotherapy, histological types and residence were observed between the two groups ($P > 0.05$) (see **Table 1**).

Immune Function Indexes

All immune function indexes were notably higher in EG than in CG ($P < 0.001$), as presented in **Table 2**.

Negative Emotions

The HAMA and HAMD scores in EG were remarkably lower compared with CG ($P < 0.05$) (see **Table 3**).

Serological Indicators

The VEGF and MMP-2 levels were markedly lower in EG than in CG ($P < 0.001$) (**Figure 1**).

Post-traumatic Growth Levels

The PTGI score in EG was notably higher when compared with CG ($P < 0.001$), as shown in **Figure 2**.

QOL

The scores of physiological function, psychological function, social function and physical function in EG were obviously higher compared with CG ($P < 0.001$), as shown in **Table 4**.

DISCUSSION

Lung cancer (LC), the most common cause of cancer death, is also an important contributor to the damage and great economic pressure (17). Though playing an important role in the clinical treatment of LC, radiotherapy and chemotherapy trigger serious side effects, such as symptoms such as fatigue, cough, shortness of breath, and decreased appetite after chemotherapy, resulting in a serious decline in QOL of patients (18). Therefore, how to reduce or alleviate the side effects of chemotherapy and enhance the QOL of patients has become an urgent problem to be solved. With the insufficient understanding of the disease, LC patients are often accompanied by serious negative emotions such as annoyance, fear and depression with the clinical manifestations of distress, reduced self-esteem and crying once they are informed of the disease (19). While seriously disturbing the

TABLE 1 | Comparison of clinical data.

Items	EG	CG	X^2/t	*P*
Gender			0.047	0.829
Male/Female	28/17	27/18		
Average age (Mean ± SD, yrs)	55.67 ± 4.65	55.24 ± 5.00	0.422	0.674
BMI (Mean ± SD, kg/m^2)	20.64 ± 1.72	20.52 ± 1.56	0.347	0.730
Times of chemotherapy				
The first time	13 (28.89)	14 (31.11)	0.053	0.818
The second time	14 (31.11)	15 (33.33)	0.051	0.822
The third time	12 (26.67)	11 (24.44)	0.058	0.809
Fourth or above	6 (13.33)	5 (11.11)	0.104	0.748
Histological types				
Squamous cell carcinoma	22 (48.89)	21 (46.67)	0.045	0.833
Adenocarcinoma	14 (31.11)	17 (37.78)	0.443	0.506
Small cell carcinoma	6 (13.33)	5 (11.11)	0.104	0.748
Large cell carcinoma	3 (6.67)	2 (4.44)	0.212	0.645
Family income			0.179	0.673
>3,000 Yuan/(month·person)	25 (55.56)	23 (51.11)		
≤3,000 Yuan/ (month·person)	20 (44.44)	22 (48.89)		
Residence [*n* (%)]			0.403	0.525
Urban area	22 (48.89)	19 (42.22)		
Rural area	23 (51.11)	26 (57.78)		
Education [*n* (%)]				
College degree and above	4 (8.89)	6 (13.33)	0.450	0.502
High school	13 (28.89)	15 (33.33)	0.207	0.649
Junior high school and below	28 (62.22)	24 (53.33)	0.729	0.393

TABLE 2 | Comparison of immune function indexes [Mean ± SD].

Group	*n*	IgG	IgA	IgM
EG	45	8.96 ± 0.95	6.69 ± 0.58	4.56 ± 0.25
CG	45	6.02 ± 0.49	3.15 ± 0.51	1.32 ± 0.11
t		18.450	30.747	79.576
P		<0.001	<0.001	<0.001

clinical treatment, these negative emotions also affect the sleep and QOL of patients. Supported by the Internet and based on the positive psychology, the network-based positive psychological nursing mode aims to alleviate the negative emotions of patients undergoing chemotherapy through the implementation of the positive psychological intervention, so as to explore their advantages, help them establish the confidence and courage to actively face the disease, alleviate negative emotions, and boost the QOL. A recent study (20) has confirmed that cognitive-behavioral therapy in network groups can effectively improve the anxiety and depression of LC patients after chemotherapy, and help patients to face treatment with a positive attitude. However, this study only discusses the psychological emotions of patients, and lacks in-depth discussions on their immune function and QOL. Elemene injection is an anti-cancer drug developed independently in China. With certain curative effects on a variety of malignant tumors, it can enhance the immune function of patients, and interfere with the synthesis of DNA and RNA in tumor cells accompanied by less effects on normal cells, which has been confirmed in tumor diseases such as malignant pleural effusion, advanced gastric cancer and esophageal cancer (21, 22). This paper discussed the effect of the network-based positive psychological nursing model combined with elemene injection on LC patients undergoing chemotherapy, and analyzed the immune indexes and serological indexes of patients, so as to clarify the clinical application value. Based on the above method, this experiment selected 90 LC patients treated with chemotherapy in our hospital as the research subjects, in which the EG group received the network-based positive psychological nursing model combined with elemene injection to analyze the effect of different treatment methods on the patients and clarify the clinical application value by comparing the immune function indexes, serological indicators and other indexes after treatment between the two groups. As shown in **Table 1**, no obvious

TABLE 3 | Comparison of negative emotion scores [Mean ± SD].

Group	*n*	HAMA score	HAMD score
EG	45	15.33 ± 4.28	19.33 ± 5.96
CG	45	19.36 ± 5.27	23.82 ± 4.51
t		3.982	4.030
P		<0.05	<0.05

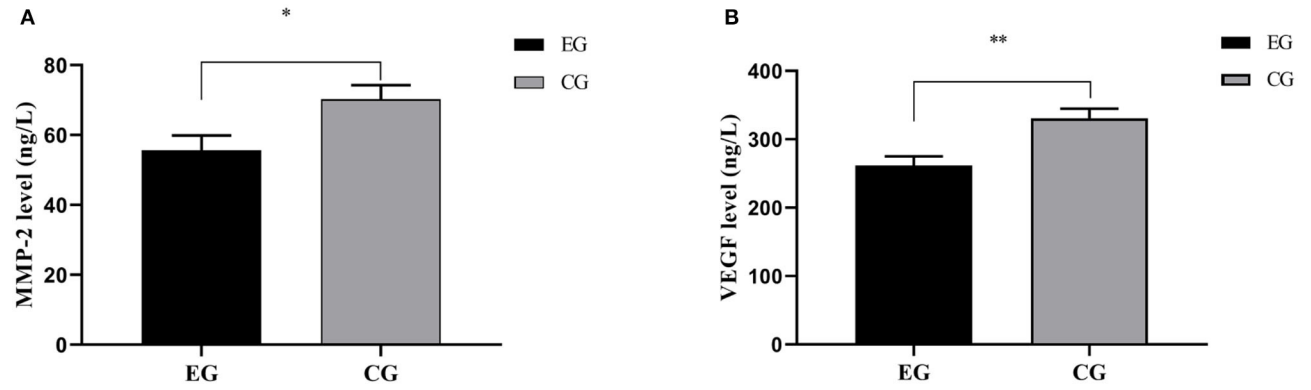


FIGURE 1 | Comparison of serological indicators [Mean ± SD]. **(A)** Comparison of MMP-2 levels. The abscissa represented EG and CG, and the ordinate represented the MMP-2 level (ng/L). The MMP-2 levels of EG and CG were (55.61 ± 4.28) ng/L and (70.27 ± 4.06) ng/L, respectively. *Indicated an obvious difference in the MMP-2 levels between the two groups ($t = 16.670$, $P < 0.001$). **(B)** Comparison of VEGF levels. The abscissa represented EG and CG, and the ordinate represented the VEGF level (ng/L). The VEGF levels of EG and CG were (261.86 ± 13.29) ng/L and (330.83 ± 14.09) ng/L, respectively. **Indicated an obvious difference in the VEGF levels between the two groups ($t = 23.887$, $P < 0.001$).

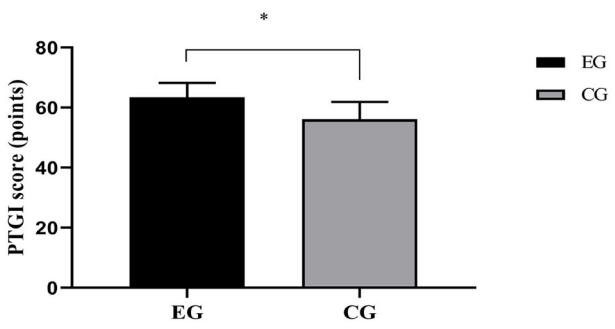


FIGURE 2 | Comparison of PTGI scores [Mean ± SD]. The abscissa represented EG and CG, and the ordinate represented the PTGI score (points). The PTGI scores of EG and CG were (63.42 ± 4.78) and (56.18 ± 5.69), respectively. *indicated a notable difference in the PTGI scores between the two groups ($t = 6.536$, $P < 0.001$).

differences in age, BMI, times of chemotherapy and histological types were observed between the two groups ($P > 0.05$), which met the requirements of clinical trials.

Under pathological conditions, activated MMP-2 can degrade the extracellular matrix components and break through the matrix barrier, leading to tumor invasion and metastasis. VEGF is a highly specific angiogenic growth factor to promote the formation of new blood vessels, which maintains a high level in many tumor diseases (23). This study showed that the VEGF and MMP-2 levels after treatment were obviously lower in EG than in CG ($P < 0.001$), suggesting that this therapeutic schedule can significantly reduce the VEGF and MMP-2 levels in LC patients undergoing chemotherapy, possibly because elemene injection with the effect of anti-angiogenesis can effectively inhibit the proliferation, invasion and migration of tumor cells (24). In terms of the improvement of negative emotions, the HAMA and HAMD scores in EG were markedly lower compared with CG ($P < 0.05$). This is because the network-based positive psychological nursing model takes individual negative psychological level as the research goal and adopts positive emotions to resist negative emotions, so as to stimulate personal potential, tap personal advantages, help patients actively face the setbacks and misfortunes, and urge them to experience the joy and success experience brought by positive actions. Therefore, their self-worth and sense of identity are enhanced to fundamentally improve the negative emotions. As for immune function and QOL, they were notably better in EG than in CG after intervention ($P < 0.001$). The reason may be that elemene injection is metabolized through the lungs and has a high concentration in lung tissues with less side effects and good tolerance in patients, which can effectively prolong the survival time and enhance the QOL of patients. These effects of elemene injection have been confirmed in the third-line treatment of advanced non-small-cell lung cancer (25). In addition, an animal experiment (26) has shown that elemene injection can improve the T lymphocyte functions of mice, and

TABLE 4 | Comparison of QOL scores [Mean ± SD].

Group	*n*	Physiological function	Psychological function	Social function	Physical function
EG	45	71.96 ± 3.75	75.44 ± 4.27	76.80 ± 4.64	77.60 ± 4.82
CG	45	67.91 ± 3.36	64.93 ± 3.14	67.82 ± 3.89	70.18 ± 4.74
t		5.396	13.302	9.949	7.363
P		<0.001	<0.001	<0.001	<0.001

enhance the antigen presentation and expression of immune-related factors. The network-based positive psychological nursing model pushes network courses about the positive psychological intervention to patients via mobile APPs, and encourages them to write gratitude logs to imperceptibly cultivate their optimistic attitude toward life, transfer the pain caused by adverse reactions of chemotherapy, and boost the QOL. Contributions: After the network-based positive psychological nursing model combined with elemene injection was applied to LC patients undergoing chemotherapy, this study was aimed to effectively alleviate the negative emotions of such patients, reduce the impact of chemotherapy on their immune function, and enhance QOL by evaluating the clinical indicators of patients after treatment. This nursing model also provides an important basis for improving the treatment confidence osf cancer patients, reflects great progress and development of clinical nursing and has comprehensive guiding significance in modern health education and clinical practice, which will undoubtedly become the future direction of medicine. However, psychological nursing based on positive psychology has not yet formed a unified operation mode, which needs to be continuously expanded and deepened. In addition, its localization research under the cultural background in China needs to be further strengthened.

CONCLUSION

In conclusion, the network-based positive psychological nursing model combined with elemene injection is a reliable method to enhance the immune function and QOL of LC patients undergoing chemotherapy. This intervention is beneficial to patients and is superior to routine nursing intervention in improving patients' negative emotions, reducing their serological indicators and enhancing their QOL, with significant effects.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Jimo Nanquan Health Center. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SY, LZ, and ZL: conception and design, data analysis, and interpretation. All authors contributed in administrative support, provision of study materials or patients, collection, assembly of data, manuscript writing, and final approval of manuscript.

REFERENCES

1. Hamai K, Fujitaka K, Kitaguchi S, Tamamoto S, Takao S, Nishida A, et al. Gemcitabine maintenance therapy after gemcitabine and platinum drug chemotherapy for naive stage IIIB/IV squamous cell lung cancer: a phase II study. *Anticancer Drugs.* (2021) 32:767–72. doi: 10.1097/CAD.0000000000001079
2. Igawa S, Naoki K, Shintani Y, Sekine I, Shukuya T, Takayama K, et al. Survival and prognostic factors in elderly patients receiving second-line chemotherapy for relapsed small-cell lung cancer: Results from the Japanese Joint Committee of Lung Cancer Registry. *Lung Cancer.* (2020) 146:160–4. doi: 10.1016/j.lungcan.2020.05.038
3. Shinde A, Horne ZD Li R, Glaser S, Massarelli E, Koczywas M, et al. Optimal adjuvant therapy in clinically N2 non-small cell lung cancer patients undergoing neoadjuvant chemotherapy and surgery: The importance of pathological response and lymph node ratio. *Lung Cancer.* (2019) 133:136–43. doi: 10.1016/j.lungcan.2019.05.020
4. Cortellini A, Maio MD, Nigro O, Leonetti A, Cortinovis DL, Aerts JG, et al. Differential influence of antibiotic therapy and other medications on oncological outcomes of patients with non-small cell lung cancer treated with first-line pembrolizumab versus cytotoxic chemotherapy. *J Immunother Cancer.* (2021) 9:e002421. doi: 10.1136/jitc-2021-002421
5. Wang TT Li YH, Zhang M, Wang SM, Zhao AG, Chu L. et al. Development of improved version of Quality of Life Assessment Instrument for Lung Cancer Patients Based on Traditional Chinese Medicine (QLASTCM-Lu). *Chin J Integr Med.* (2019) 25:831–6. doi: 10.1007/s11655-018-2991-5
6. Hyland KA, Small BJ, Gray JE, Chiappori A, Creelan BC, Tanvetyanon T, et al. Loneliness as a mediator of the relationship of social cognitive variables with depressive symptoms and quality of life in lung cancer patients beginning treatment. *Psychooncology.* (2019) 28:1234–42. doi: 10.1002/pon.5072
7. Chan JC, Lee YH, Liu CY, Shih HH, Tsay PK, Tang WR, et al. Correlational study of skin toxicity and quality of life in patients with advanced lung cancer receiving targeted therapy. *J Nurs Res.* (2019) 27:e51. doi: 10.1097/jnr.0000000000000339

8. Child ST, Walsemann KM, Kaczynski AT, Fleischer NL, McLain AC, Moore S. Personal network characteristics and body mass index: the role of education among Black Americans. *J Public Health (Oxf).* (2019) 41:130–7. doi: 10.1093/pubmed/fdy020
9. Limoges J, Nielsen K, MacMaster L, Kontni R. Globally networked learning: deepening Canadian and Danish nursing students' understanding of nursing, culture and health. *Nurse Educ Today.* (2019) 76:228–33. doi: 10.1016/j.nedt.2019.02.006
10. Alizada M, Li J, Aslami H, Yang D, Korchuganova T, Xu YH, et al. viaβ-Elemene Inhibits the Proliferation and Migration of Human Glioblastoma Cell Lines Suppressing Ring Finger Protein 135. *Balkan J Med Genet.* (2020) 23:43–9. doi: 10.2478/bjmg-2020-0002
11. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. (2013) 27:2191–4. doi: 10.1001/jama.2013.281053
12. Wang Y, Jing MX, Jiang L, Jia YF, Ying E, Cao H, et al. Does a ketogenic diet as an adjuvant therapy for drug treatment enhance chemotherapy sensitivity and reduce target lesions in patients with locally recurrent or metastatic Her-2-negative breast cancer? Study protocol for a randomized controlled trial. *Trials.* (2020) 21:487. doi: 10.1186/s13063-020-04429-5
13. Borges K, Campos CS, Amaral AC, Camila M, Coelho DP, Neto AEH, et al. Nursing documentation for chemotherapy in a university hospital's bone marrow transplant unit: a best practice implementation project. *Int J Evid Based Healthc.* (2020) 18:75–85. doi: 10.1097/XEB.0000000000000216
14. Kaboli PJ, Salimian F, Aghapour S, Xiang S, Zhao Q, Li M, et al. Akt-targeted therapy as a promising strategy to overcome drug resistance in breast cancer - A comprehensive review from chemotherapy to immunotherapy. *Pharmacol Res.* (2020) 156:104806. doi: 10.1016/j.phrs.2020.104806
15. Plana AL, Nogueira PF, Guardiola PM, Sánchez SS, Roda EM, Montoyo HP, et al. The novel proautophagy anticancer drug ABTL0812 potentiates chemotherapy in adenocarcinoma and squamous nonsmall cell lung cancer. *Int J Cancer.* (2020) 147:1163–79. doi: 10.1002/ijc.32865
16. Sun F, Franks K, Murray L, Lilley J, Wheller B, Banfill K, et al. Cardiovascular mortality and morbidity following radical radiotherapy for lung cancer: Is cardiovascular death under-reported? *Lung Cancer.* (2020) 146:1–5. doi: 10.1016/j.lungcan.2020.05.004
17. Yang XR, Man JY, Chen H, Zhang TC, Yin XL, He QF, et al. Temporal trends of the lung cancer mortality attributable to smoking from 1990 to 2017: a global, regional and national analysis. *Lung Cancer.* (2021) 152:49–57. doi: 10.1016/j.lungcan.2020.12.007
18. Kyte K, Ekstedt M, Rustoen T, Oksholm T. Longing to get back on track: patients' experiences and supportive care needs after lung cancer surgery. *J Clin Nurs.* (2019) 28:1546–54.
19. Yu J, Huang T, Xu J, Xiao J, Chen Q, Zhang LX. Effect of nursing method of psychological intervention combined with health education on lung cancer patients undergoing chemotherapy. *J Healthc Eng.* (2022) 2022:2438612. eCollection 2022 doi: 10.1155/2022/2438612
20. Noro R, Kobayashi K, Usuki J, Yomota M, Nishitsuji M, Shimokawa T, et al. Bevacizumab plus chemotherapy in nonsquamous non-small cell lung cancer patients with malignant pleural effusion uncontrolled by tube drainage or pleurodesis: a phase II study North East Japan Study group trial NEJ013B. *Thorac Cancer.* (2020) 11:1876–84. doi: 10.1111/1759-7714.13472
21. Wang B, Song K, Chen LH, Su HY, Gao LY, Liu JF, et al. Targeted inhibition of ACK1 can inhibit the proliferation of hepatocellular carcinoma cells through the PTEN/AKT/mTOR pathway. *Cell Biochem Funct.* (2020) 38:642–50. doi: 10.1002/cbf.3522
22. Jabari M, Farsani MA, Salari S, Hamidpour M, Amiri V, Mohammadi MH. Hypoxia-Inducible Factor1-? (HIF1α) and Vascular Endothelial Growth Factor-A (VEGF-A) expression in de novo AML Patients. *Asian Pac J Cancer Prev.* (2019) 20:705–10. doi: 10.31557/APJCP.2019.20.3.705
23. Lin L, Wang QY, Xu F, Luo XL, Xu J, Yan LP, et al. BML-111, the lipoxin A agonist, modulates VEGF or CoCl-induced migration, angiogenesis and permeability in tumor-derived endothelial cells. *Immunol Lett.* (2021) 230:27–35. doi: 10.1016/j.imlet.2020.12.007
24. Wang YH, Shi XH Qi QH, Ye B, Zou ZL. Safety of anlotinib capsules combined with PD-1 inhibitor camrelizumab in the third-line treatment of advanced non-small-cell lung cancer and their effect on serum tumor markers. *J Healthc Eng.* (2021) 2021:2338800. doi: 10.1155/2021/2338800
25. Ogura Y, Kataoka N, Kunimatsu Y, Tachibana Y, Sugimoto T, Tani N, et al. Predictors of survival among Japanese patients receiving first-line chemoimmunotherapy for advanced non-small cell lung cancer. *Thorac Cancer.* (2021) 12:97–105. doi: 10.1111/1759-7714.13720
26. Ma LT, Bai Y, Li J, Qiao Y, Liu Y, Zheng J. Elemene emulsion injection administration reduces neuropathic pain by inhibiting astrocytic NDRG2 expression within spinal dorsal horn. *Chin J Integr Med.* (2021) 27:912–8. doi: 10.1007/s11655-021-3438-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Zheng, Sun and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 10 May 2022
doi: 10.3389/fpubh.2022.897565

Research on Diagnosis and Management of Postgraduates Mental Health Status Based on BP Neural Network

Jiangning Xie[1,2]*

[1] School of Management, Shandong University, Jinan, China, [2] Graduate School, Shandong University, Jinan, China

As a special group facing the huge pressure of study, employment, life, and other aspects, graduate students have had fierce psychological conflicts. Their mental health status has been greatly affected. Postgraduate mental health problems are related to students' normal life and study and the scientific research work during graduate study. Because of this, this paper uses the BP neural network to establish a diagnostic model for postgraduate mental health status with the dependent variables of SCL-90 psychometric test results to realize the diagnosis of postgraduate' mental health status. The accuracy of the training, validation, and test process of the BP neural network is 98.448, 97.373, 98.128%, and the overall fitting accuracy of the model is 98.273% and the prediction accuracy of graduate mental health is 97.93%. Finally, according to the mental health status of graduate students to ensure the smooth progress of scientific research during graduate students.

Keywords: BP neural network, postgraduate mental health, diagnosis, management, graduate students

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Srividhya S. R.,
Sathyabama Institute of Science and Technology, India
Kuruva Lakshmanna,
VIT University, India
Priya N.,
Jain University, India

***Correspondence:**
Jiangning Xie
xjn@sdu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *16 March 2022*
Accepted: *04 April 2022*
Published: *10 May 2022*

Citation:
Xie J (2022) Research on Diagnosis and Management of Postgraduates Mental Health Status Based on BP Neural Network. Front. Public Health 10:897565. doi: 10.3389/fpubh.2022.897565

INTRODUCTION

The importance of postgraduate mental health is self-evident, whether to the postgraduate or the school, society, and the country. Nowadays, the pace of study, life, and postgraduate scientific research are accelerating. In addition, it is difficult to find a job after graduation, and the psychological pressure of students is increasing rapidly. In many aspects, such as daily study and scientific research, interpersonal communication, emotional life, depression, anxiety, inferiority, and other different psychological problems, which have caused a significant impact on the growth of graduate students. Therefore, attaching great importance to the mental health education of graduate students has become a critical task in graduate student training in various universities.

Research on the Mental Health of Graduate Students

Since the late 1980s, there have been more and more studies on the mental health problems of college students at home and abroad. However, most of these studies focused on the mental health status and related factors of undergraduate students, while relatively few studies on the quality of mental health in the graduate population. Research on graduate mental health can be roughly divided into two aspects. The first aspect is the causes, performance, and countermeasures of graduate student mental health problems. Ma et al. (1) investigated the mental health status and life pressure of 340 graduate students in a university. The SCL-90 results showed that the detection rate of psychological problems in graduate students was 44.16%, which explains the

prevalence of mental health problems in graduate students. Shanmuganandapala (2) also noted that nearly 25 percent of graduate students have mental health problems, such as symptoms of depression, anxiety, or seasonal affective disorders. Posselt (3) constructed a hierarchical clustering model, studied the mental health problems of nearly 20,888 graduate students in 69 universities and compared the prevalence of depression and anxiety in different research fields. The results showed that graduate students majoring in humanities, arts, and architecture have a significantly higher probability of depression and anxiety than graduate students in other majors. On the psychological status of the Chinese Academy of Chinese Medical Sciences, Wang et al. (4) found that the psychological problems of graduate students were coercion (14.02%), interpersonal disorder (7.32%), depression (8.54%), and anxiety (5.49%).

Research on Management Strategies for Graduate Mental Health Problems

Given the mental health problems of graduate students, many scholars put forward the corresponding management strategies. Among them, Huang et al. (5) discussed the impact of group counseling guided by positive psychology theory on the mental health level of graduate students. The research results show that group counseling based on positive psychology theory can improve the mental health status of graduate students, reduce their negative symptoms and perceptual stress level, and improve their self-acceptance degree. Based on the positive psychological thought, from the perspective of psychological health promotion, Liang et al. (6) constructed a four-level promotion model of graduate mental health that is based on the needs survey of graduate mental health, classification needs as the goal, promotion mode as the core, and effect evaluation as the feedback means. The results showed that good results were achieved.

To sum up, the current graduate mental health problems are mainly from the causes of mental health, the current situation of mental health education, and countermeasures. Most of these studies are theoretical, and few experts quantify the graduate mental health problems and predict their mental health problems through the relevance of the data. Because of this, this paper takes the mental health status of graduate students as the primary research line, obtains the pressure capacity of graduate students in the face of various factors through questionnaire survey, and obtains the mental health evaluation score of graduate students through the SCL-90 questionnaire. Finally, the BP neural network model with the former as the input variable and the latter as the output variable is used to diagnose and manage graduate mental health status.

QUESTIONNAIRE AND THE MODEL

Acquisition of Graduate Mental Health Data Based on Questionnaires

Questionnaire Design

For the analysis of the factors affecting graduate mental health, this paper from the social, family, school, and individual level summarizes the following 15 influencing factors such as social adaptation, social temptation, employment, future choice, new environment, economy, competition, research, interpersonal communication, scientific research, others, parents expectations, family background, character defects, conflict of interest. According to the 15 factors that affect graduate psychology, the questionnaire is formulated as follows (**Table 1** is the main content of the questionnaire):

Self-Testing of Graduate Mental Health Based on the Self-Rated Scale of the SCL-90 Test

Symptoms Self-evaluation Scale SCL-90 (7–9) is one of the world's most famous psychological test scales. It is widely used in consultation clinics or hospitals to understand the psychological situation of the patients or patients. It can also evaluate the efficacy of the evolution of the condition before and after consultation. This scale consists of 90 items, including feeling, emotion, thinking, consciousness, behavior, living habits, interpersonal relationship, diet, and sleep, and nine factors are used to reflect nine aspects of psychological symptoms. The S-Assessment Scale SCL-90 judged the mental health status.

The Neural Network Model

An neural network model consists of some nodes/neurons, set at multiple layers: the input layer, one or more hidden layers and the output layer. Each node/neuron has an activation function, which calculates how much neuron is "stimulated" (10, 11). At each layer, the collections of nodes/neurons transform the input parameters; these parameters are distributed to the next layer, which is described as (1), (2) and (3):

$$z_j^n = \sum (w_{ji}^{(1)} x^{n-1} i + w_{j0}^{(1)}) \tag{1}$$

$$a_1^n = \sum (w_{ij}^{(2)} z^{n-1} i + w_{10}^{(2)}) \tag{2}$$

$$y_1^n = F(a_1^n) \tag{3}$$

Where x represents the input to the first layer; z represents the first layer's output; i, j represents the neural network node index; $w_{ji}{}^{(l)}$ represents the weight between the jth node in the ith layer and the ith node in the $(i + 1)$th layer; $F(a_i^n)$ represents the output value of ith node in $(n + 1)$th layer after being activated by the activation function; w and w_0 represent the weight and bias between the neurons, which measures the significance of the data passed along the link (synapse). $F(a)$ employs the activation function, which employes the hidden layer's aggregated output to calculate output y.

The initial weights and biases randomly assigned, the training process continues until the desired output is obtained, which is evaluated by the cost function (4):

$$Ew = \frac{1}{2} \sum_{k=1}^{K} y(x_k, w_k) - t_k^2 \tag{4}$$

TABLE 1 | Main contents of the questionnaire.

	Influence factor	Importance and self-assessment					
1	With the rapid development of today's society, the conflict between modern ideas and traditional ideas, and the differences between positive ideological education and popular secular views, make it difficult for you to adapt to society. For the pressure to adapt to society, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
2	In today's social society, harmful social phenomena such as value loss and moral decline, and irrational thoughts and behaviors such as hedonism, money worship, and extreme individualism affect others all the time. Do you feel about the pressure of resisting social temptation?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
3	With the introduction of competition mechanisms, enterprises have higher and higher requirements for the first degree of graduate students. Facing the pressure of employment, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
4	In the face of the reform of the personnel system of social institutions, society has put forward more stringent employment standards to graduate students. With the expansion of graduate student enrollment, the gold content of master's graduate students has become lower and lower, and the employment pressure is becoming greater and greater. Many students face difficult choices in doctoral study and employment. Do you feel about the pressure on this choice?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
5	Entering the environment, everything needs to adapt again, leaving their parents to live alone. Some students even leave their city to come to a strange town, facing the pressure caused by such environmental changes, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
6	With the rising prices and the continuous expenditure of various expenses during the graduate years, do you feel about the economic pressure?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
7	Graduate students are talents selected by the university. They are all talents with excellent academic performance in the class. After coming to graduate school, they come to a new competition stage, and almost everyone has the same starting point of competition. In the face of this competitive pressure, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
8	Facing the triple pressure of graduate scientific research, study, and tutor, do you feel too burdened? Secondly, there is a graduate school, the examination of various qualification certificates, etc. In the face of this learning pressure, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
9	The students come from different regions, with different political and cultural backgrounds and personal habits, especially those in the dormitory. Due to regional differences, there will inevitably be friction in the process of getting along. In the face of these interpersonal pressures, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
10	The postgraduate period is the enlightenment stage of scientific research. Much scientific research work is not conducted during the university. Do you feel about the pressure of facing new knowledge and new fields?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
11	For students with excellent performance and outstanding performance, it is easy to get various awards from the school, such as scholarships, exceptional student honor, etc., which will inevitably be envied by other students and thoughtfully calculated by others. In the face of these pressures, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
12	For graduate students, they are all the pride of their parents. Many parents let their children prepare to go abroad, study a doctoral study and seek a high salary job in advance. Are they facing the pressure of their parents' strong expectations?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
13	Each student has a different family background. Those students with ordinary family conditions hope to graduate as soon as possible, work hard, and earn money to support the family. Some families even give up the idea of studying because of financial problems. Do you feel about the pressure of family economic conditions?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee

(Continued)

TABLE 1 | Continued

	Influence factor	Importance and self-assessment					
14	People will inevitably have personality, cowardice, inferiority, withdrawn, narrow, impulsive, irritable, self-centered nature will be reflected in the individual. This kind of personality is often challenging to get along with others. Do you feel about the interpersonal pressure caused by personality defects?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee
15	There is inevitable friction between students, but many students tend to take themselves as the center, think that others have to consider themselves, especially when they cannot meet their expectations in the psychological imbalance, in the face of the pressure caused by this conflict of interest, do you feel?	Score	A	B	C	D	E
		Evaluation	Giant	Big	Normal	Small	Wee

Data source: Independent writing.

TABLE 2 | Questionnaire results.

	Question															SCL-90
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	3	5	5	1	4	4	4	3	5	2	3	1	1	4	3	160
2	2	5	2	5	2	4	1	3	5	1	4	5	1	2	4	163.2
3	1	4	4	1	4	5	2	3	5	4	5	2	1	5	2	156.8
4	5	5	4	5	3	1	1	1	3	1	5	1	1	1	5	137.6
5	2	4	5	5	5	5	4	3	2	5	4	4	2	3	1	179.2
457	1	2	1	1	2	2	1	1	5	3	2	4	3	5	1	121.6
458	4	1	4	1	5	5	1	2	2	3	5	1	3	2	4	147.2
459	1	4	1	2	5	2	3	4	1	2	5	5	5	5	5	176
460	5	2	1	5	5	5	1	5	1	4	3	3	2	4	3	169.6
461	5	1	3	4	4	5	1	2	3	1	1	4	5	4	5	156.8

Where y represents the output; t represents the desired output, w represent the weight, Levenberg–Marquardt (LM) algorithm is utilized in the neural network training process, which is a variation of gradient descent. The weight and bias of the neural network model are changed during the training process to minimize the error, which is described as function (5):

$$w^n = w^{n-1} - \left(J^T J + \mu I\right)^{-1} J e^{n-1} \tag{5}$$

Where $J=\frac{\partial E}{\partial w}$represents the full-scale Jacobian matrix related to w; I represents the identity matrix; m represents a combination coefficient; e represents the prediction error.

The Levenberg–Marquardt algorithm starts with a forward computation by (1), (2) and (3). The prediction errors of the output layer and the hidden layer are calculated by (6), (7) and (8):

$$e^{(3)} = y_1 - t \tag{6}$$

$$\delta_1^3 = e_1^3 \tag{7}$$

$$\delta_j^2 = w_{1j}\delta_1^3 \tag{8}$$

As shown in function (9) and function (10), the Jacobian is calculated by a back-propagation process:

$$\frac{\partial E}{\partial w_{ji}} = \delta_j^{(2)} x_i \tag{9}$$

$$\frac{\partial E}{\partial w_{1j}} = \delta_1^3 z_j \tag{10}$$

In the training process of the sample, the learning sample should be processed to make it fluctuate in a certain range. The normalization method is adopted in this paper to process the data to ensure that the data is between 0 and 1, which is written as (11):

$$x_i = \frac{\bar{x}_i - x_{\min}}{x_{\max} - x_{\min}} \tag{11}$$

Where $\bar{x}_i$ represents the average value, $x_{\max}$ represents the maximum value, $x_{\min}$ represents the minimum value.

TABLE 3 | Mental health diagnosis results.

	Actual value	Predicted value	Error (%)		Actual value	Predicted value	Error (%)		Actual value	Predicted value	Error (%)
1	144	147.39	2.35	21	173	171.96	0.60	41	138	134.09	2.83
2	166	169.03	1.83	22	166	169.69	2.22	42	199	189.31	4.87
3	137	139.13	1.55	23	150	148.47	1.02	43	128	134.39	4.99
4	128	127.76	0.19	24	150	152.02	1.35	44	166	156.36	5.81
5	153	156.02	1.97	25	157	158.69	1.08	45	186	183.33	1.44
6	144	143.76	0.17	26	179	178.89	0.06	46	144	142.74	0.87
7	109	116.74	7.10	27	128	129.56	1.22	47	150	149.02	0.65
8	141	141.41	0.29	28	125	125.25	0.20	48	170	173.86	2.27
9	186	180.32	3.05	29	131	131.75	0.57	49	150	146.26	2.49
10	150	152.67	1.78	30	144	143.57	0.30	50	166	173.00	4.22
11	147	148.95	1.33	31	137	132.62	3.20	51	109	114.98	5.49
12	176	180.84	2.75	32	140	137.61	1.71	52	157	160.44	2.19
13	147	145.86	0.78	33	160	156.76	2.03	53	134	131.43	1.92
14	156	149.26	4.32	34	144	141.52	1.72	54	147	152.50	3.74
15	163	164.89	1.16	35	134	136.92	2.18	55	147	144.64	1.61
16	144	142.70	0.90	36	173	176.90	2.25	56	186	181.03	2.67
17	144	145.24	0.86	37	157	160.79	2.41	57	122	126.51	3.70
18	150	153.17	2.11	38	131	127.68	2.53	58	147	147.90	0.61
19	172	168.76	1.88	39	173	178.54	3.20	59	176	179.24	1.84
20	131	129.98	0.78	40	147	144.49	1.71	60	169	166.95	1.21
								61	157	153.39	2.30
Average error (%)						2.07					
Overall accuracy (%)						97.93					

Data source: Author organizes.

RESULTS AND DISCUSSION

Questionnaire Results Statistics

The respondents of this questionnaire are 500 graduate students coming from different schools and majors. Five hundred questionnaires were sent out, 487 were recovered, and 461 were valid questionnaires, with an efficiency of 92.2%. For the 15 questions in the questionnaire, options A-E correspond to 1–5 points, respectively, and a higher total score indicates more significant psychological stress. The questions 1–15 and the corresponding SCL90 scores are shown in **Table 2** (partial results).

Attention

According to the score standard of SCL90, the total score is more than 160, should be further examination, the standard is more than 200 points that there are obvious psychological problems, can turn to psychological counseling, more than 250 is more serious, need a detailed medical examination, is likely to do targeted psychological treatment or take medicine under the guidance of a doctor.

Graduate Student Mental Health Diagnosis Model Establishment

Fifteen factors affecting graduate psychology were used as independent variables based on the above questionnaire. The corresponding SCL-90 mental health diagnosis scores were dependent variables to establish the BP neural network model. The data for the training model has randomly selected 400 sets, with 80% training set, 20% test set, and 20% validation set. The remaining 61 sets of data were control groups, and the error between the SCL-90 score and the actual SCL-90 score was analyzed to verify the model's utility (12).

The training procedure of this model is shown in **Figure 1**.

The training process of the BP neural network model is shown in **Figure 1A** and improves the above neural network configuration. The regression model is initialized in the training process, with the ideal error convergence rate in step 40. **Figure 1B** shows that high fitting accuracy is obtained after the training process. Then, the accuracy of the training, validation, and test process is 98.448, 97.373, 98.128%, and the overall fitting accuracy of the model is 98.273%. The model's accuracy is high, especially in the test, where the predicted values of the output parameters are close to the reference value. Results show that the relationship between the input and output parameters is modeled correctly. Further prediction/estimation is reliable. Moreover, the error histogram of the model is shown in **Figure 1C**. For all models, 20 bins are distributed according to Gaussian rules, proving the representativeness of the training data.

Application of the Graduate Mental Health Diagnostic Model

By establishing the BP neural network model, the mental health of the remaining 61 students was diagnosed, and the questionnaire results of these 61 students were input into the prediction model to obtain the following output results and

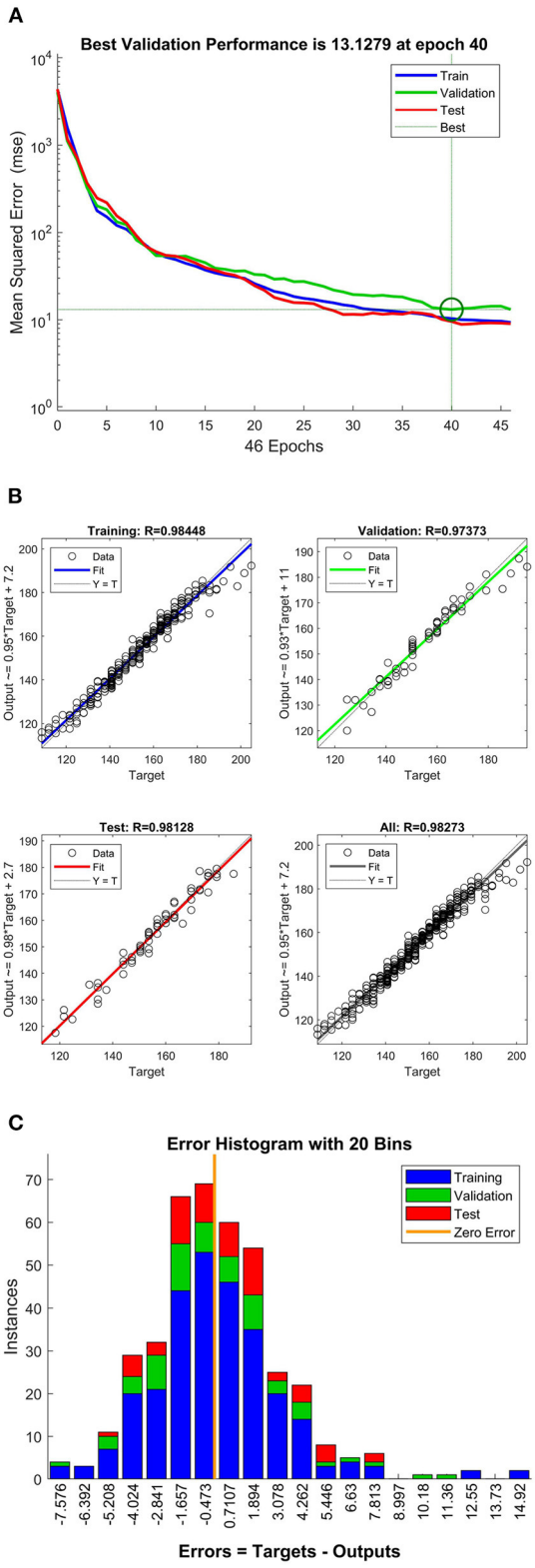


FIGURE 1 | Training process of the model; **(A)** convergence characteristics; **(B)** regression performance of the model; **(C)** error histogram of the model.

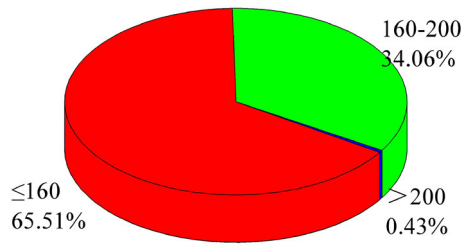


FIGURE 2 | Statistics of mental health scores for graduate students. Data source: Author organizes.

compare the output results to the actual results of SCL-90, as shown in **Table 3**.

As shown in **Table 2**, the trained diagnosis model achieved an excellent diagnostic effect in practical application. The overall accuracy reached 97.93%, an average error of 2.07%. Among them, the largest diagnostic error of 7.10%, through the above analysis, it can be seen that the mental health diagnosis model in practice or has very high practical value, the judgment of students with psychological problems is more accurate, achieved the expected effect.

Management of Graduate Mental Health Problems

Figure 2 shows the mental health diagnosis scores of the 461 graduate students in this survey. As shown in **Figure 2**, 34.06% of graduate students had mild psychological problems in the survey results, and 0.43% had obvious psychological problems. Still, they could be solved through psychological counseling. No students in this survey scored more than 250 points, indicating that the surveyed graduate students could solve their psychological problems through positive guidance methods such as psychological counseling.

It can be seen from this survey that the psychological problems of graduate students are common, but they can be solved through positive guidance.

Given the above analysis, this paper proposes the management strategy of graduate psychological problems:

(1) Improve the tutor responsibility system

During graduate school, the relationship between tutor and graduate student is the most important interpersonal relationship. In the training process of graduate students, the communication between tutors and graduate students is the closest. Attention should be paid to building an excellent psychological contract between "mentor and graduate students." Colleges and universities should improve the tutor responsibility system, give full play to the guiding role of tutors, and clarify the importance of tutors for enhancing the comprehensive ability of graduate students. Graduate tutors should always be strict with themselves in teachers' ethics, set an example in scientific research work, and abide by the principle of integrity.

The tutor has a subtle influence on students 'mental health problems in scientific research and life. The tutor's guidance

and evaluation will directly affect the self-efficacy of graduate students. Sometimes the tutor's sentence can greatly encourage students' enthusiasm for scientific research. When students encounter scientific research problems, they should give timely guidance. When students have mastered the experimental techniques and principles in the field, they should lead students to study the significant cutting-edge problems in the area appropriately.

Due to the expansion of graduate enrollment, the number of tutors increases, so it is difficult to consider all graduate students. In order to improve the quality of graduate training, the relevant departments can appropriately control the number of students. According to relevant surveys, some tutors use the "free-range" mode for graduate students and meet less often, so it is difficult to timely understand the students' scientific research status in time, it is difficult for students to get reasonable guidance, and it is easy to cause too much pressure on scientific research. So suggest tutor according to their own and team, appropriate increase for graduate research guidance and humanistic care time, can be through group form, once a week or once every 2 weeks, timely understand everyone's scientific research and life situation, give full play to the mentor lead and research guidance, create a good relationship between teachers and students, reasonable reduce and avoid graduate mental health level, reduce or relieve the pressure of medical graduate research.

Due to the expansion of graduate enrollment, the number of tutors increases, so it is difficult to consider all graduate students. To improve the quality of graduate training, the relevant departments can appropriately control the number of students. According to relevant surveys, some tutors use the "free-range" mode for graduate students and meet less often. So it isn't easy to understand the students' scientific research status in time timely. It is difficult for students to get appropriate guidance, and it is easy to cause too much pressure on scientific research. Therefore, according to their situation and the research group, the tutors can appropriately increase the time of scientific research guidance and humanistic care for graduate students. They can timely understand everyone's scientific research and life situation and give full play to the role of the tutors in the guidance and scientific research guidance. Finally, it will create a good teacher-student relationship, reasonably reduce and avoid the low level of mental health of graduate students, and reduce or relieve the scientific research pressure of medical graduate students.

(2) Strengthen the mental health education in colleges and universities

Graduate mental health education is crucial in the graduate training stage. We should start from the actual situation of mental health education in colleges and universities to strengthen the mental health education of graduate students and effectively help postgraduate students to relieve psychological pressure. Colleges and universities can establish mental health counseling rooms or online consultation topics by psychological doctors. Graduate students can timely solve their problems in life and scientific research and prevent them from deteriorating into psychological issues.

(3) Increase the scientific research support for graduate students

The generation of scientific research results is inseparable from the support of universities for scientific research work. Therefore, to reduce the pressure of students in scientific research, the schools need to provide vital support in the foundation of scientific research. According to the research subject direction of the institute, each department buys perfect public scientific research facilities so that students can obtain scientific research results through experiments and reduce their burden on scientific research.

(4) Counselors actively carry out psychological counseling work

College counselors are a high-quality team with strong combat effectiveness. They should give full play to the role of counselors and strengthen the attention of graduate students. College counselors should pay attention to graduate mental health education. As a healthy vulnerable group, their mental health level is generally low. Female graduate students have delicate and rich feelings, are sensitive to any stimulation, have a changeable mood, have relatively fragile psychology, and are prone to dependence, inferiority, depression, and other psychological problems. Therefore, counselors should pay attention to the mental health of graduate students, especially the mental health of female graduate students, including family situations and love situations. Counselors can set up a mutual aid group for graduate students to fully play the exemplary role of senior teachers and senior brothers.

CONCLUSION

Under the background of graduate mental health tension, through the study of the graduate mental health factors, and based on design graduate mental health evaluation questionnaire and SCL-90 mental health evaluation, using the BP neural network model to graduate mental health diagnosis, and verify the feasibility of the diagnostic model. After the final model validation, the prediction accuracy of graduate mental health is 97.93%, which indicates that the mental health diagnostic model established in this paper can evaluate reasonably the mental health status and is feasible.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.897565/full#supplementary-material

REFERENCES

1. Ma XT, Li WH. Research on Graduate mental health status and life stress. *Spec Educ China*. (2011) 4:6. doi: 10.3969/j.issn.1007-3728.2011.04.015
2. Shanmuganandapala B, Khanlou N. An interview between a professor and a nursing graduate student on the current state of post-secondary student mental health in Ontario, Canada. *Int J Ment Health Addict*. (2019) 17:418–24. doi: 10.1007/s11469-019-00104-2
3. Posselt J. Discrimination, competitiveness, and support in US graduate student mental health. *Stud Grad Postdr Educ.* (2021) 12:89–112. doi: 10.1108/SGPE-07-2020-0042
4. Wang CY, Du XL, Di B, Kung SS, Zeng L, Peng T, et al. Investigation on mental health of graduate students of Chinese Medical Sciences and analysis of influencing factors. *Chin J Tradit Chin Med Inform*. (2016) 23:4.
5. Huang WQ, Zhang R, Liu YX, et al. The effect of group guidance on improving graduate mental health research-based on the theory of positive psychology. *Chin J Clin Psychol*. (2012) 20: 4.
6. Liang SH, Zhu YH, Zhu WE. Building a graduate mental health promotion model from the perspective of positive psychology. *J Univ Electron*. (2018) 20:5. doi: 10.14071/j.1008-8105(2017)-1159
7. Prinz U, Nutzinger DO, Schulz H, et al. The symptom-check-list-90-R (SCL-90-R) and the short versions of the SCL-90-R: psychometric analyses of inpatients with mental disorders. *Phys Med Rehabil Kurortmed.* (2008) 18:337–43. doi: 10.1055/s-0028-1093323
8. Bech P, Bille J, MLler SB, Hellström LC, Østergaard SD. Psychometric validation of the Hopkins Symptom Checklist (SCL-90) subscales for depression, anxiety, and interpersonal sensitivity. *J Affect Disord.* (2014) 160:98–103. doi: 10.1016/j.jad.2013.12.005
9. Grande TL, Newmeyer MD, Underwood LA, Williams CR. Path analysis of the SCL-90-R: exploring use in outpatient assessment. *Meas Eval Couns Dev*. (2014) 47:271–90. doi: 10.1177/0748175614538061
10. Yang X, Liu H, Peng Y. Enhanced perceptual psychopathology correlation mechanism model based on rough set and neural network. *Cluster Comput*. (2019) 22:1–7. doi: 10.1007/s10586-018-1908-6
11. Yang A, Zhuansun Y, Liu C, Li J, Zhang C. Design of intrusion detection system for internet of things based on improved BP neural network. *IEEE Access.* (2019) 7:106043–52. doi: 10.1109/ACCESS.2019.2929919
12. Yuan ZP, Zhu D. Research on the impact of scientific research on graduate mental health- -take the school of transportation engineering as an example. *Knowl Econ*. (2019) 12:2.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fpubh.2022.881124

Social Intervention and Governance of Youth School Bullying—Based on Computer Medical Data Analysis

*Jiahui Zhao**

School of Government, Shanghai University of Political Science and Law, Shanghai, China

The rise of computational social science provides a new method for campus bullying research based on large-scale data collection, calculation and analysis. Governing the bullying behavior of a middle school through social intervention, and closely observe the service needs and existing problems of the school youth group. This paper analyzes the characteristics, inducements and negative effects of school bullying. Combine drama courses and working group education methods to intervene in school bullying. Intervention work includes making teenagers aware of bullying behavior and identifying bullying types. To achieve the purpose of empathy through role play, bullies can effectively control irrational thoughts, understand their own cognitive biases, and reconcile their own emotions and behaviors. So that the victims can identify the bullying behavior around them in time, and cultivate their resistance and self-protection awareness in the event of bullying. Based on the empirical analysis of social work to intervene in the practical dilemma, and put forward the corresponding countermeasures to reduce the negative impact of school bullying on all aspects of youth, so as to reduce the various social risks brought by school bullying.

Keywords: school bullying, teenagers, big data, social work, drama education, behavioral interventions

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Chennareddy Vijay Simha Reddy,
Middlesex University, United Kingdom
Kadiyala Ramana,
Annamacharya Institute of Technology and Sciences, India

***Correspondence:**
Jiahui Zhao
yuz79906@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *22 February 2022*
Accepted: *12 April 2022*
Published: *20 May 2022*

Citation:
Zhao J (2022) Social Intervention and Governance of Youth School Bullying—Based on Computer Medical Data Analysis. Front. Public Health 10:881124. doi: 10.3389/fpubh.2022.881124

INTRODUCTION

At present, campus bullying is common. Campus bullying is a kind of behavior that intentionally injures the personal and property safety of others, including beating, intimidating, cursing, isolating, controlling, insulting, laughing and so on. Rodríguez-Muoz et al. (1) investigated the phenomenon of bullying in local primary and secondary schools. After in-depth analysis, he believes that intimidation is the main manifestation of bullying behavior among teenagers in school. Bullying is a behavior that continuously harms others. Due to the imbalance of social and physical and mental strength between the victim and the abuser, the victim suffers long-term and continuous harm in the process of bullying (1, 2).

In terms of school bullying (2), stated that bullying means that "in the same group, the powerful party uses words and behaviors to intentionally bring physical and mental stress and harm to the vulnerable group members". Japan's education ministry in the early 1990s defined bullying as "unilateral, sustained physical and mental attacks that can cause great pain to others".

Park et al. (3) points out that "Bullying is a pervasive aggressive behavior. This behavior is often characterized by threatening or intimidating others." Webb et al. (4) believes that school bullying refers to the intentional and violent assault, including physical and psychological assault, on others by means of beating, harassment, isolation and

other means. According to Ran et al. (5), campus bullying refers to the continuous bullying of the strong against the weak and the large group against the small group in the campus, which is in conflict with or directly related to the high school students (3, 4).

According to the above literature description, it is not difficult to find that although there is no unified definition of school bullying, it can directly or indirectly reflect the asymmetry of power between the bully and the bullied. Carreira et al. (6) believes that bullying is a common social phenomenon with the characteristics of concealment, diversity, repetition, imbalance, universality and indistinguishable (5). Georgiou et al. (7) believes that bullying is characterized by diverse forms, violent behavior, fragile object and extensive influence (8).

To effective governance adolescent bullying behavior, improve the adolescent bullying phenomenon, at the middle school has carried on the empirical research, this paper adopts the social work intervention combined with drama class, effectively improve the middle school campus bullying problems, and analyze the intervention of social work practice difficulties, some countermeasures are put forward. Finally, it provides references for enriching bullying-related theories. The research on campus bullying in my country started late and the pace is slow, the literature is not concentrated and the understanding is not comprehensive, and most of them are analyzed and studied from the perspective of psychology and sociology. This research adopts practical investigation, analyzes the situation and influencing factors of the problem with the help of theoretical tools, and tries to put forward strategies of management, implementation and guarantee for the prevention and treatment of school bullying from the perspective of pedagogy theoretical research.

STUDENT GROUP NEEDS AND SOCIAL WORK INTERVENTION METHODS

The Needs of Student Groups in the Context of School Bullying

In recent years, the incidents of campus violence have been frequent, which has aroused widespread concern in the society (9, 10). The Office of the Education Supervisory Commission of the State Council issued a special action plan on cracking down on school bullying in May 2018, calling for a severe crackdown on intentional intimidation and bullying by primary and middle school students using physical, verbal and online technology. In addition, a survey on the harmonious situation of teenagers in 10 middle schools in a certain area was conducted. The self-reported bullying rate was 22.3%, the self-tested bullying rate was 13.8 and 33.5% of students said they had witnessed bullying, indicating that it was more serious (11, 12). School bullies seem to be everywhere, and they have different degrees of influence on the growth of teenagers. Even the "mass psychology" of bullying leads children to use similar solutions and imitate the bully to behave in unsafe ways. It threatens the physical and mental development of primary and middle school students. On this basis, the author believes that it is necessary to teach the bully and the victim to express their thoughts from the perspective of emotional control, irrational expression and control, replace violence with correct expression, help the bully and the victim to establish a good relationship, and ultimately reduce the occurrence of bullying (13, 14).

Working at a social workstation in a middle school, I got to know the details of the middle school as follows: Grade 7 was divided into 6 classes with 260 students, Grade 8 was divided into 8 classes with 340 students, Grade 9 was divided into 8 classes with 240 students. The school provides educational resources not only for registered students, but also for children of migrant workers with excellent academic performance. So the students come from different sources (15, 16). The ratio of family enrolments to migrant students is currently about 6:4. In addition, classes will be arranged for students, including those with learning disabilities and disabilities, in a number of classrooms.

Through questionnaires and observations, social workers have found that different groups in schools will need different services by 2019–2020. Group service and demand analysis are shown in **Tables 1**, **2**.

This suggests that eighth-graders, students on loan and students with different behaviors are more likely to have school bullying (17, 18). Without further research, the author reached a consensus with the school to organize eighth grade students to participate in street campus bullying exhibition and campus bullying prevention drama working group, so as to help control campus bullying and create a harmonious campus.

Combined with the results of the unstructured interview, it is concluded that there are both bullies and victims in the working group. Most of them have four problems. Problem 1: Bullies, victims, and bystanders do not recognize that some of the actions they are dealing with are bullying, and do not know how to deal with bullying. Problem 2: At the family level, there are problems such as poor family communication, inadequate family structure, lack of care or inappropriate parenting methods, which are all factors that lead to violent behavior. Bullies often believe that their parents do not understand them, there is a lack of communication and a generation gap. The families of the victims lack communication, they don't like to talk at home, so the parents can't observe the inner changes of the children. Problem 3: In school, teachers tend to pay more attention to students who are good in grades and obedient. Bullies often think that teachers always take themselves as targets or negative teaching materials, so they don't listen to the teacher's orders, and arouse the teacher's attention and recognition by singing the opposite. Problem four: in interpersonal relationships, bullies and victims cannot find intimate friends, they are low self-esteem, lack of confidence. Cowardly and cowardly students are often bullied and ostracized; Students who bully others often hope to establish their status through violence. Bullies believe that power makes people obedient, afraid, safe, and valued. So, the members of the task force have several requirements: First, be aware of bullying. Secondly, it is necessary to understand interpersonal relationships and interpersonal communication skills. The second is to learn how to prevent school bullying and how to protect yourself when it happens (19, 20).

TABLE 1 | Group services and demand analysis.

Group	Demand
Seventh grade students	1. Environmental adaptation needs: freshmen in the seventh grade cannot adapt to the middle school life due to the changes in learning content, work and rest rules and campus environment. 2. Running-in needs for interpersonal communication: Because teenagers are in a special stage of growth and development, belonging to the running in period of self and society, their hearts are more sensitive, it is difficult to integrate into the collective and establish a stable interpersonal relationship with classmates. 3. Adolescent attention needs: Grade 7 students are in a critical period of physical and psychological development, which can also be called the rebellious period. During this period, great changes occur in body and mind, which requires more attention from parents and teachers.
Eighth grade students	1. Running-in needs of interpersonal communication: Grade 8 students have not left the growth and development period of teenagers, so it is difficult for them to integrate into the group and establish a stable interpersonal relationship with their classmates, and some bullying behaviors may occur. 2. Adolescent Guidance Needs: Grade 8 students have adapted to their physical and mental changes during adolescence. However, due to their ignorance of the opposite sex, they tend to have puppy love or stay away from the opposite sex. 3. Learning objective planning needs: Due to the physical and mental changes, the eighth grade students' attention is easy to be shifted, it is difficult to focus on learning, lack of learning goals and learning motivation, so they can not keep up with the curriculum, resulting in inferiority complex.
Ninth grade students	Planning of learning direction: Grade 9 students have initially formed an independent personality. Due to the change of puberty, students are focusing on interpersonal relationship, which leads to their loss of learning goals and motivation, and they have no specific plan for future study.

TABLE 2 | Special group services and needs analysis.

Group	Demand
Students	1. Adapting Habits: As the student still has the habits brought by the former campus and interpersonal environment, these habits may be in conflict with the habits of the students in the current campus, such as hygiene habits, work and rest rules, dialect, etc., which hinders the student's integration and adaptation in the new environment. 2. The need to enhance family attention: due to the changes of study and interpersonal environment, students are more sensitive and prone to inferiority, so parents need to pay more attention to the changes. 3. The need for acceptance among classmates: due to the changes in the new environment, students hope to make new friends as soon as possible and keep up with the new pace of study, so they need to be accepted by classmates.
Poor students	1. Increase the need for self-confidence: due to the lack of economic conditions, poor students are very likely to have a sense of inferiority, or extreme conceit. 2. Material needs: Compared with other types of student groups, poor students lack material conditions more. They need basic material conditions from their families to reduce the gap between them and their classmates.
Students with different behaviors	1. Constraint and guidance needs: Because parents and teachers do not pay attention to students' different behaviors, students do not know that some of their behaviors are wrong, resulting in students' cognitive and behavioral bias. 2. The need to be tolerated: because their individual behaviors are not understood, other students will label them, and even discriminate against and misinterpret them. Therefore, such students need to be treated with tolerance.

Purpose and Method of Social Work Intervention

Recruit bullies and victims of bullying and use professional social work tools to intervene in bullying and violence incidents. To understand the psychological needs of students, give appropriate treatment, so as to achieve better results, improve the extreme psychology of bullies, so that they reduce aggressive behavior. Develop the victim's confidence in the outside world so that he or she can seek help when confronted with bullying problems (21, 22). A good campus environment helps to correct students' personal behavior, so the construction of a "harmonious campus" environment can enable students to receive good education, so that students have normal interpersonal relations, conducive to the overall development of body and mind.

The use of social work intervention to deal with campus bullying is the main way of social work. The form of social work intervention can promote the working group and its members to achieve their needs, so that students can accelerate their own social transformation through working group. Coordinate and develop interpersonal relationships between individuals and working groups; So as to play the social functions of organizations and groups, improve the social psychology of bullying and victims, so that they can get a sense of social identity and belonging. After establishing a clear goal, team members can help each other and collectively obtain more resources. It makes the members of the working group actively communicate, understand each other, have the courage to take responsibility and get treatment in the mutual cooperation, so as

to promote positive changes in the attitude and behavior of the team members (23, 24).

The working group always adopts the way of role play. Role-playing is to understand a variety of behavior habits and norms in the society by playing the role of others, learn to understand the psychological activities and behavior content of others, and finally realize the social transformation of team members. By playing different interpersonal roles, we can gain insight into the positions and psychological activities of the roles we play. In the process of empathy, we can get a better way to deal with campus bullying and interpersonal relationship.

Drama courses are also the basis for social work interventions. Through the drama course, the working group realized the interpretation and correction of school bullying. Drama courses should not only learn the knowledge and interpretation skills of drama, but also enable students to understand the society, think about life, control behavior and protect themselves through drama. Drama course is to introduce drama activities into the field of education and use drama background to achieve educational purposes.

The purpose of drama course in social work intervention is to change the common phenomenon of campus bullying. So that the bully can understand the hurt of the bully through scene modeling, role playing and other methods, and make the bully brave to ask for help when hurt.

ANALYSIS OF SCHOOL BULLYING BEHAVIOR

Characteristics of School Bullying Behavior

Bullying is not an occasional argument or fight. It is an unfair behavior where one party constantly puts pressure on the other. Whether the bystander, the bully or the bullied, campus bullying will bring profound physical and mental harm. In recent years, there have been media reports of "a dispute case concerning the right to health accepted by a certain court". Some junior high school girls beat up another girl, leaving her handicapped. The incident started when the victim did not pay attention to personal hygiene during school, which caused hostility from other female students in the dormitory. Five schoolgirls slapped her in the face to "educate" her, resulting in a grade 10 disability.

After finishing the junior high school Chinese test, 16-year-old Xiao Huang endured great pain and continued to take the senior high school entrance examination. For 4 years, he kept a secret from his parents: Starting in the fifth grade, he was regularly caught in "unprovoked attacks" by several other students. This kind of school bullying is not uncommon in the media.

A simple message can tell us that bullying is insidious, persistent and communal. School violence is frequent and hidden. Bullying often takes place out of the reach of teachers and parents. For example, dormitory and toilet, whose daily use is relatively private, it is hard to be found by others. Many victims have been bullied for years but do not have the courage to ask their parents and teachers for help. This cowardice makes the bully more aggressive, less moral, and more stubborn. School bullying has a fixed pattern, that is, victims are always bullied; Bullying is not an individual act. It occurs in small groups and groups. They have similar characters and imitate each other in their actions. Bullying behavior is a group behavior formed by a person under multiple individual pressures.

Causes of School Bullying

Middle school students are in a critical stage of physical and mental development, they often do not have enough experience to deal with problems harmoniously, and they cannot empathize with others. So the way they deal with some of these problems is to emulate other students. Personality traits also play an important role in bullying behavior. Bullies generally have strong personalities and need external recognition and "face saving". They believe that bullying can make them stronger and save face more. Bullies are introverted and timid. They dare not ask for help when being bullied. They can only swallow their pride and become the objects of long-term bullying. There are also some victims of bullying who have birth defects or bad habits. They are often mocked and insulted, but their innate condition makes them unable to resist.

Bullying is not only a personal problem for students, family relationships are also an important cause of school bullying. Every family has a different way of raising children. Some parents do not care about their children's life and psychological state, lack of guidance and education for children's growth, resulting in the lack of correct supervision and guidance of children's behavior, it is easy to lead to bullying behavior; In some families there is tension and parents often scold their children. This leads to a lack of security and the belief that violence is the only way to solve the problem. Lead to bullying behavior to solve conflicts with peers; In some families, the parents are very bossy and don't care about the inner needs of the children. They only demand obedience and obedience from the children. The child who receives this kind of family education will become weak, self-abased, unsociable, etc., and the child will often become the object of bullying. Therefore, the family relationship and parents' education style will decide the child's growth performance (25, 26).

The school environment, as the cradle of children's education and growth, is another factor that leads to bullying behavior. The strict management mechanism of the school will greatly restrict the bad behavior of the school, so that good communication between teachers and students is not only the responsibility of teachers, but also an important condition for students to grow up. To enhance teachers' sense of mission in education, make them pay attention to students' academic performance and psychological state, so as to reduce students' bullying behavior. Or when bullying behavior occurs, teachers can perform their duties, effectively curb the continuation of bullying behavior, communicate with parents in time, and do a good job in family work between the bully and the bullied, so as to solve the problem of bullying on campus from the root.

The Negative Effects of School Bullying

The primary problem of school bullying is that the victim is physically bullied. These include beating, shoving and other acts that may cause physical injury to the victim. Because trauma is especially easy to spot, it's also easy for the media to see physical

bullying. Because bullies fear that the outside world will find out about their bad behavior, they choose to beat their victims by avoiding obvious places, such as the face and limbs. Most of the victims have wounds in the abdomen and lower body, leaving internal wounds that are hard to see outside. The health of the victims has been greatly affected.

Moreover, the victims have suffered irreparable psychological trauma. Bullies have been abused, beaten and teased for a long time. Because of the imbalance of power, he will endure hidden psychological pressure and pain for a long time, which will lead to the formation of a pessimistic personality, negative, cowardly and fearful character, not only difficult to maintain their studies, they often fear in life, do not know how to live, do not know how to get along with others. Some victims of intolerance turn to extreme behaviors, such as suicide or self-harm, which can have a devastating effect on their mental health and development. Bullies, on the other hand, usually have poor grades and are "face-saving", and they use violence to prove their authority. In addition to forming "cliques" among their peers, they also have difficulty communicating with others and often threaten vulnerable groups with violence. It is easy to break the law and crime after becoming an adult, which has a negative impact on social development.

SOCIAL WORK INTERVENTION

Working Group Intervention Objectives

The overall goal of the task force's intervention was to reduce the irrational thinking and behavior of bullying participants through drama lessons and role-playing, to develop a more comprehensive understanding of bullying, and to identify appropriate skills to deal with bullying. Establish good interpersonal relationship and promote campus harmony.

The phased objectives of the intervention by the working group are as follows: firstly, to improve the students' overall awareness of school bullying, so that the working group members can understand the concept, types, hazards and manifestations of school bullying. The second is to change the irrational thinking of bullies, break through the communication barriers between bullies and teachers, parents and classmates, learn interpersonal communication skills, and establish a good interpersonal relationship. The third is to enable the victims of bullying in the group to learn not to let others hurt them, let the bullies know that they have the ability to control their own aggressive behavior, and promote them to establish the values of helping others and harmonious progress together.

Working Group Intervention

Working Group Name: "Peace Elite" Campus Drama Curriculum Working Group

Intervention period: 2:00–4:00 p.m. every Wednesday from March to June 2020

Number of activity classes: 6

Duration: 120 min per session

Group size: 6 people

(1) Section 1: Establishment of working group relationship. See **Table 3** for details.

The first drama course is the first step in the development of social work, marking the initial stage of social work intervention in the treatment of school bullying. Since the working group has just been set up and the members are not familiar with each other, the social workers will choose sand tools and sand paintings to let the members show themselves and get to know each other, and to create an atmosphere of mutual trust and ease in the group. Understand and interact with team members on their work goals and expectations to increase their participation and sense of belonging.

Course summary: After the social worker introduced himself first, some team members were gradually led to introduce themselves, but some team members did not adapt to the team atmosphere, and their performance was relatively cold. Working groups should start with ice-breaking sessions so that members can interact more and have more opportunities to get to know each other. The members of the later working group have reached the stage of mutual understanding, and have a certain impetus to the working group, from which the cohesion gradually formed. After that, the communication and motivation between team members will be further strengthened through role playing.

(2) Section 2: How to identify school bullying. See **Table 4** for details.

Course Analysis: As the second stage of social work intervention, this course solves the problem of members' lack of participation and unity in the early stage of working group construction through the ice breaking link. At the same time, in order to make the team members know more about campus bullying, the working group should also be responsible for organizing the interaction and questioning among the team members, and explaining the relationship between drama course and campus bullying for the team members. It is also necessary for the members of the group of social work organizations to watch the film, share their impressions and encourage them to express their views through sharing. And in the process of watching the film, the freeze frame screen specifically analyzes the specific types of campus bullying, so that the students participating in the working group can better understand the knowledge of campus bullying. The second section of the course reduces the resistance of group members to the intervention of social work, and the internal vitality of the working group increases with the deepening of the course, gradually forming cohesion (27, 28).

Course Summary: Compared with the first working group course, in the process of carrying out the second course, the team members have eliminated the formality and constraints. Therefore, compared with the first course, this course has been recognized by more team members, and the interaction within the group is gradually frequent. At the beginning of the course, participants were not very active, but when the social worker shared the meeting, participants were attracted; When watching the video, I was very focused. After watching the film, the group members can enthusiastically discuss its content. So the social worker must personally instruct the team members,

TABLE 3 | Working group activities and establishment of group relations.

Number of sessions	Working group objective/working group activities content	Material
Section I	Working group goal: the team members are familiar with each other, understand each other, establish the working group goal. 1. Introduction to staff and members 2. Use sand table and sand painting activities to organize group members' internal interaction and improve group members' sense of identity. 3. Determine the activity content and stage goal of the working group, and obtain the expected effect of the team members on the working group. 4. Establish the contract of the working group and the implementation standard of the activity 5. Notice in the next section.	Sand table, white paper, marker pen, sand ware

TABLE 4 | Campus bullying identification course.

Number of sections	Working group objective/working group activities content	Material
Section 2	Working group goal: To enable all team members to identify what constitutes bullying and to distinguish the types of bullying. 1. The warm-up game: Who's undercover 2. Discuss in pairs, put forward what you think about school bullying, and share your understanding of drama courses. 3. Play the school bullying movie "Teenage You", share your impressions and introduce the bullying knowledge in it. 4. Pause some scenes in the movie, and analyze the type characteristics of such bullying behaviors. 5. Summary speech and the next section notice.	Stool

prepare more cases to share with them, and use multimedia to attract them into the team. Second, members of the working group responded strongly to the video and could easily ask or answer other less relevant questions during the discussion. In this lesson, due to the divergent thinking of middle school students, it is necessary to bring the subject of the course back to the issue of school bullying in time when the course is off topic.

(3) Section 3 Working Group: To recognize their own irrational thoughts. See **Table 5** for details.

Course Analysis: The working group has entered the third session on intervention in school bullying. In this lesson, the focus is to understand the irrational thinking of the bully. To understand what irrational thinking is, let the group members find irrational thinking in their daily life. Further strengthening the contact and communication between the groups provides more opportunities for communication and interaction between the groups, and also enables the bullies in the group to learn to solve the problem in a more reasonable way. Bullies reflect on their daily behaviors and thoughts, so that group members can have a deeper understanding of themselves, learn how to deal with irrational thinking, figure out how to control and reduce irrational thinking, and learn how to rebuild rational thinking in group intervention. To fundamentally solve the extremist thoughts of the bully to lay a certain foundation of the course.

Lesson Summary: Bullies in groups tend to be more actively involved in thinking and acting, and they have difficulty staying silent during interactions with others. Bullies like to "rhythm" and interfere with others' sharing, and tend to stray from the group's theme. Social workers need to remind and emphasize the working group's curriculum themes, as well as discipline issues in the curriculum. Because social workers need to spend a lot of time and energy on maintaining the theme and order, their ability to grasp the course schedule is relatively weak. In order to consolidate and practice the daily communication skills and methods of the working group in the teaching of the course, before the end of the course, the social workers assigned some assignments on the subject of the course, and shared their experience in the next class, so as to ensure that the team members understand the consequences of irrational thinking.

(4) Section 4 Working Group: Emotion Control and Interpersonal Communication Skills. See **Table 6** for details.

Course Analysis: The fourth part of the working group continues to focus on the issue of emotional control for bullies. The bullies in the group shifted from the resistance group to the acceptance and trust group. In this lesson, the task force shares several ways to calm down a bully in an irrational mood. In addition, through role-playing skills, the team members simulated the psychological activities they experienced in different environments, used "I" to express

TABLE 5 | The unrational behavior courses for understanding the bullying self.

Number of sessions	Working group objective/working group activities content	Material
Section 3 (Bullies)	Task group goal: Make the bully aware of his irrational thoughts 1. Ask the bully in the group to share the irrational thoughts that they have on a daily basis and explain why. 2. Use the self-questioning method to ask the school bully some of his own irrational thoughts, do I have to do this? If I was injured, what should I do? And discuss new solutions with everyone. 3. Through some irrational behaviors in real life, this paper analyzes the thinking process of self rhetorical question through cases, and then puts forward a reasonable approach through discussion. 4. Make a summary speech and encourage the bullying participants to make a difference. 5. In the next section.	Work notes

TABLE 6 | Emotional control and interpersonal skills.

Number of sessions	Working group objective/working group activities content	Material
Section IV (Bullies)	Working Group's goal: to enable school bullies to effectively control their emotions and acquire interpersonal skills. 1. In the emotional more excited, through effective skills to gradually calm down. 2. Role-playing is adopted to set up a specific social environment so that the bully can play a student who can solve his own emotional problems. 3. Express yourself in soothing, gentle, and polite language by introducing your ideas. 4. Summarize the speech and affirm the change of the bully's behavior.	Work notes, water pens

their thoughts every day, and strengthened their control over bad emotions. Let the bully represent the bullied, learn to put yourself in others' shoes, so as to achieve the thinking change of "putting yourself in others' shoes", understand the harm of school bullying behavior, and reject bullying behavior from the heart.

Course Summary: The content of the working group is very complete. In this course, the group members have learned how to calm down their emotions, get rid of the control of irrational emotions, experience the inner changes in the role so as to achieve the requirement of self-reflection. Bullies can actively participate in the interaction when they play other roles. Although some students with good face avoid interaction because they are afraid of making a fool of themselves, they finally complete the role playing through the persistence of social workers and the encouragement of their peers. Through communication and interaction, the team members' trust in the team and the friendship among them are enhanced. The group work is progressing smoothly, but there is a lack of negative emotion venting and mature emotion control system in the course. Therefore, we need to adopt appropriate incentive policies and reasonable material rewards, so that team members can more actively participate in the interactive activities of the working group.

(5) Section V Working Group: Behavior Attribution of Bullies. See **Table 7** for details.

Course Analysis: The subject of this course is the victim of bullying. First, ask the victim to share positive things in their life, as well as brave things in their daily life. According to the characteristics of the victims, the thinking mode of the victims when they encounter school bullying is analyzed from different perspectives. The teaching video "Know Yourself" and "Snack Fight" were shown. The victims were asked to describe the personality and mood of each character in the video, as well as the influence of their way of doing things. Share the bullying behavior and its characteristics in the video, so that they can get the opportunity to understand themselves in the communication, and lay the foundation for changing the bullying.

Lesson Summary: When the victim joins the work group, the group members have no emotional involvement and are only attracted to the group when they watch a short film shown by the work group. The group members were relatively quiet and could not relax during the activity. Social workers must play ice-breaking games at the beginning of classes, as well as more interactive sessions. Make the team members who have been bullied have confidence in the working group, and more conducive to enhance the cohesion of the group.

(6) Section 6 Working Group: How to say no to school bullying in terms of behavior. See **Table 8** for details.

Course Analysis: This course is the last course of social work intervention, and also the last stage of school bullying management. The last class of the working group went smoothly and recorded the school bullying play performed by the whole group by means of photography and video. Members of the team create mementos by recording the scripts they create, produce and execute themselves into video clips. At the end of the stage play, the social workers arranged all the members to write a letter from the heart, put the letter in a time capsule and buried it under the tree on campus. Because of the active participation in the group, the team members unite, trust and encourage each other in the course, and finally learn to correctly understand campus bullying, how to reasonably control emotions and protect themselves. Finally, the staff

TABLE 7 | Behavior lying of bullied behavior.

Number of sessions	Working group objective/working group activities content	Material
Section V (Bullying)	Working group goals: To understand the thoughts of the bullied, analyze the bully's situation and behavioral attributions. 1. The victim shares their positive energy of being brave in life. 2. Games: the right interpretation. 3. Examples are given to illustrate some thinking patterns and behavioral logics in the individual environment of different team members. 4. Conduct behavioral attribution questionnaire and boundary perception expression 5. Play teaching videos "Know Yourself" and "Battle for Snacks" to analyze the time, roles, emotions, ideas and influences in the videos. 6. How can I protect myself after sharing lessons. 7. Summarize the speech and encourage the victims to fight back and seek help.	Work notes, water pens

TABLE 8 | Say no to bullying on campus.

Number of sections	Working group objective/working group activities content	Material
Section 6 (All team members)	Working Group Goal: Show how the bully's real thoughts can be expressed properly through role play, and how the bullied can say no in the face of school bullying. 1. Theater stage performance. 2. Each team member writes a letter from his heart, puts it in a time capsule, and buries it in the campus together. 3. The staff members share and summarize the changes of the team members in the activity. 4. Take photos to commemorate. 5. Finish the job.	Camera, camera, work notes, time capsule model

shared the changes of all team members in this course, and encouraged and recognized all the students who participated in the working group.

Summary: The working group has entered the final stage of intervening in school bullying. At this stage, the members have been able to independently recognize their responsibilities and unite with each other. Compared with the initial stage of the group, the group members' identity with the group and with each other has changed remarkably. During this period, the group members had a deeper understanding of school bullying in the development of group courses, and they also recognized the teaching methods of drama courses more. In the process of sharing and interacting with each class, social workers actively guide students' thinking direction and respond to students' questions. Actively guide the students to summarize their own learning results, so that they can flexibly use the interpersonal skills learned in the drama course in real life. In this for the intervention of social work, draw a perfect full stop.

Drama Course Effect Evaluation

At the end of the drama course, students participating in the drama course were asked to fill in the drama course effect self-scale, and the impact of drama course on school bullying was evaluated according to the proportion of students in the drama course. The results are shown in **Table 9**.

As shown in the table above, In the self-assessment, most of the team members are very clear about the meaning of campus bullying, and have a good understanding of the skills and knowledge of drama course. Through the drama course, the team members have a better understanding of the thinking process of the bully and the bullied, and learned to think in other places, which can improve the bullying behavior on campus. Thirty percent of the students think that they can solve problems with their own methods, which shows that the knowledge taught by the working group cannot be easily used in practice. But more than half of the team members clearly realized that violence did not solve the problem. And realize that violence is a kind of behavior that brings negative influence to the students around them. Only one member passed the scale with full marks. He thought that he could better understand himself through this group. It can also make other team members realize some of their own ideas and understand others' positions and behaviors by themselves and others. Most of the group members don't want to encourage their friends to join the working group, but they are happy to participate in other activities of the social workstation.

SOCIAL WORK INTERVENES IN PRACTICAL DIFFICULTIES

Diversity of Types of School Bullying

There is a wide variety of types of bullying and bullying behavior is not fixed. School bullying can be divided into three main categories: the first is verbal bullying. Verbal bullying is a common phenomenon in daily life on campus. Abuse, personal attacks, nicknames and so on are all verbal bullying; The second category is behavioral bullying. Behavioral bullying

TABLE 9 | Drama course effect self—scaletitle.

	Number of people	Proportion (%)
I can identify the types of bullying on campus.	6	100
I learned about drama classes	6	100
I can control my irrational thoughts	5	83.3
I've mastered interpersonal skills	6	100
I learned what to do when there is bullying on campus	6	100
I can handle my ideas in the right and reasonable way	4	66.6
I can lead by example and be a peace elite	5	83.3
I learned that escape or violence is not the right solution	6	100
I understand how unreasonable some of the past behavior, I will control or protect themselves.	6	100
I think drama course can effectively control campus bullying	4	66.6
I would like to continue to participate in the activities organized by the social working group	5	83.3

means that the bully intimidate others for a long time and cause physical harm to others, such as beating, deliberately pushing, tripping, punching and kicking, or even intentionally harming others' economic interests and life and property safety. Use blackmail, isolation, defamation, and coercion to do something against one's will. The third category is cyberbullying. With the progress of science and technology, the Internet has become an indispensable "necessity" for people, and the phenomenon of bullying comes along with the Internet. There is less pressure and constraint in online communication, so teenagers are more likely to vent their negative emotions without any cost when using platforms such as online social media, such as maliciously slander others and make rumors. Even in school forums, post-bars and other environments, there is bullying both online and offline. Therefore, it can be seen that the role types of school bullying are diverse.

Personal Cognitive Biases of School Bullying

Students' personal factors play an important role in the prevalence of campus bullying. The cognition between people directly affects people's behavior. From the perspective of cognitive behavior theory, among the three factors of emotion, cognition and behavior, cognition plays a role of neutralization and coordination in emotion and behavior. Understanding and observing the same things in many ways will lead to different behaviors. Different cognitive biases of the same behavior can also lead to this result. In terms of campus bullying, with the deepening of bullying, the bully becomes more and more bossy and arrogant; However, the bullied become more and more cowardly. Some people fight against Campus bullying, while others choose to avoid self comfort. People's different views on campus bullying are also due to cognitive bias. The bully usually pursues power and face, thinks that he can get the experience of maximizing ability in bullying, and pursues security and sense of belonging through personal authority; Those who are bullied are often timid, dare not face the bullying, dare not speak up. Lead to their own long-term threat and injury, and ultimately that resistance is useless. Obviously, these are caused by personal cognitive bias.

When social workers asked students, "What is bullying in your mind?", the students will only one-sided answer: being beaten, coercion. Most students don't realize that verbal abuse, teasing and isolation are also bullying. In order to better solve and prevent the problem of school bullying, schools need to train students to have a correct understanding of school bullying. The task force's research includes the recognition of irrational thoughts and attribution of thought and behavioral outcomes, the use of therapy to correct emotions, change false cognitive biases, and thus transform the way school bullies and victims think and behave, and develop interpersonal skills and confidence. To empower young people to speak up and reject bullying.

The Emergency Management Mechanism of School Bullying Is Not Perfect

As a social worker, the author went to a high school and proposed to set up a school bullying working group. However, after the school learned of the author's intention, the school did not agree to set up educational courses and working groups on the subject of "school bullying". Later, after the author found social workers, the school decided to change the theme of the working group to "harmonious campus". During the drama class of school bullying working group, the author observed the students in each class and found the occurrence of campus bullying. There was a slight bullying incident on campus, which was only handled by the teaching director alone. When the bullying on campus is serious, the head teacher will contact the parents to give oral education. The results show that schools have no ability to control campus bullying due to the imperfect emergency management system.

SUGGESTIONS ON COUNTERMEASURES FOR SCHOOL BULLYING

Improve Teenagers' Perception of Bullying

Teenagers are in the special period of adolescence, which is the key stage of psychological development. Many times, teenagers' misunderstanding of things will lead to cognitive bias. If young people want to understand things without one sidedness, they must first understand themselves correctly and form a correct outlook on life and values, which is manifested in dignity, self-confidence and love. First, teenagers must know how to deal with interpersonal relationships, learn interpersonal skills,

respect others, especially in middle school life, and learn to control irrational thoughts. When there is conflict with others, we should be able to deal with the problem rationally. We should not use violence and bullying to solve the contradiction. Second, strengthen students' consciousness of self-protection; When they are bullied, they know how to actively resist and take the right way to protect their own interests and security.

Establish a Good Family Relationship and Educational Environment

Parents are children's first teachers, parents' behavior will affect the children's life. The difference of family relationship and family education mode has an imperceptible effect on the development of adolescent personality. Parents should set an example and guide and educate their children to abide by laws, regulations and moral norms. They should not take a rude, non-interference and intolerant attitude toward the difficulties and setbacks their children encounter in the process of growth. Be patient, responsible and careful to teach them. At the same time, family members should communicate with each other at any time to create a harmonious and equal family atmosphere and family relations, promote family harmony and enhance parent-child relationship. Parents must keep close contact with schools and social workers, keep an eye on the phenomenon of school bullying, understand their children's behavior and life in school, track their children's psychological state and correct their psychological differences, so as to eliminate school violence. It is necessary to correctly guide and educate teenagers in the way of interpersonal communication, enhance their awareness of self-protection and prevent the occurrence of school bullying.

Improve the Emergency Management Mechanism of School Bullying

In the case of school bullying, schools must take the main responsibility. In order to prevent the occurrence of campus bullying, the campus should take the following measures to manage bullying incidents. First, the campus should use billboards, broadcasts, publicity boards, forums, official websites and other media to create a campus atmosphere of mutual assistance, friendship, respect for teachers and education. The second is to set up the legal education and mental health class for middle school students, so as to popularize the legal knowledge and establish the legal awareness, so that students can know, abide by and understand the law, do not use violence to threaten the safety of other people's lives and property, and do not use bullying to coerce others into harmful behaviors. To cultivate students' good psychological quality, learn to release bad emotions and pressure, and educate students to reduce bullying behavior on campus. Third, teachers should pay more attention to the phenomenon of school bullying. The phenomenon of school bullying is mostly caused by teachers' neglect of students and students' incomprehension of teachers. Therefore, teachers should be fair to every student and can't ignore them. Do not despise or discriminate against school bullying, maintain sensitive contact, timely intervention to the school and parents, education, care and protection of students. Fourth, schools must cooperate with social workers to hold regular class meetings to discuss the problem of school violence. For students who have all kinds of bullying behaviors on campus, social workers should guide them to follow up and observe or intervene, and use professional theoretical knowledge and skills to solve the problem of school bullying, so as to change their thoughts and behaviors in an all-round way.

Construction of Social Workstations and Multiple Network Mutual Assistance Mechanisms

To reduce the incidence of school bullying, social workstations should combine school and network to establish multiple mutual assistance mechanisms. Through the linkage and mutual assistance of "social work, school, family, network, and community", we should give full play to the advantages of all aspects and mobilize all forces to protect the victims of bullying to the maximum extent and curb the phenomenon of school bullying. Through the school, family, network, community and other platforms, provide advanced services, use professional social work methods, distribute school bullying brochures at the school gate, to promote and identify school bullying education; To prevent and respond to bullying, schools should work with social workers to provide students with regular training on how to deal with bullying on campus. To clarify teachers' responsibilities and pay more attention to students' physical and mental state, so as to provide strength for the elimination of school bullying; Social workers must maintain contact with parents and guide parents in educating and protecting children within the family, properly dealing with bullying and avoiding bullying behavior (29, 30).

CONCLUSION

At present, the phenomenon of campus bullying is becoming more and more serious. School bullying has become an important problem faced by schools and society. Therefore, the education department and scholars are very concerned about the problem of campus bullying, and the state has constantly issued laws and regulations to combat campus bullying, which can prevent the spread of bullying to a certain extent. As a social worker, we have the obligation to use professional theory, professional methods and social work practice skills to intervene in the core of campus bullying, and to intervene and control campus bullying from the root.

As for the problem of campus bullying, this paper combines social work intervention with drama therapy theory. The purpose is to change teenagers' cognition of campus bullying, make teenagers really understand the root of campus bullying, and improve teenagers' ability to deal with campus bullying. At the same time, for the bullies, the drama class carried out by the working group has improved their self-confidence and communication skills; For bullies, the key is to learn to control bad emotions and change irrational thoughts. The essence of the work of the working group is not only to improve the self-development of the group members, but also to pay attention to

the bullying behavior inside and outside the campus to prevent its spread. However, due to the lack of theoretical basis and practical experience of social work, campus bullying has not been completely and effectively solved. The efficiency of the working group on school bullying needs to be improved.

Therefore, it is very important to combat campus bullying from all aspects. This is not only from the perspective of social work intervention, but also from the perspective of school, family and government, so that young people have a good and healthy learning and growth environment. To lay a solid foundation for the construction of harmonious campus and harmonious society.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JZ conceived the research, designed, and drafted the manuscript.

REFERENCES

1. Rodríguez-Muoz A, Antino M, León-Pérez JM, Ruiz-Zorrilla Blanco P. Workplace bullying, emotional exhaustion, and partner social undermining: a weekly diary study. *J Interpers Viol*. (2020) 37:088626052093303. doi: 10.1177/0886260520933031
2. Ruch W, Stahlmann AG. Toward a dynamic model of Gelotophobia: social support, workplace bullying and stress are connected with diverging trajectories of life and job satisfaction among Gelotophobes. *Curr Psychol*. (2020) 1–13. doi: 10.1007/s12144-020-01046-y
3. Park I, Gong J, Lyons GL, Hirota T, Takahashi M, Kim B, et al. Prevalence of and factors associated with school bullying in students with autism spectrum disorder: a cross-cultural meta-analysis. *Yonsei Med J*. (2020) 61:909–22. doi: 10.3349/ymj.2020.61.11.909
4. Webb L, Clary LK, Johnson RM, Mendelson T. Electronic and school bullying victimization by race/ethnicity and sexual minority status in a nationally representative adolescent sample. *J Adolesc Health*. (2020) 68:378–84. doi: 10.1016/j.jadohealth.2020.05.042
5. Ran H, Le C, He X, Jiang L, Wang T, Yang R, et al. Resilience mediates the association between school bullying victimization and self-harm in Chinese adolescents. *J Affect Disord*. (2020) 277:115–20. doi: 10.1016/j.jad.2020.07.136
6. Carreira T, André C, Candido A, Santiago C, Cruz O, Amendoeira J. O3? The school as a personal construction space: how to prevent bullying? *Eur J Public Health*. (2019). doi: 10.1093/eurpub/ckz094.002
7. Georgiou SN, Charalambous K, Stavrinides P. Mindfulness, impulsivity, and moral disengagement as parameters of bullying and victimization at school. *Aggress Behav*. (2019) 46:107–15. doi: 10.1002/ab.21876
8. Kyoung Y, Kim JY, Maleku A, Moon SS. Typologies of peer victimization, depression, and alcohol use among high school youth in the united states: measuring gender differences. *Soc Work Public Health*. (2019) 34:293–306. doi: 10.1080/19371918.2019.1606750
9. Fraguas D, Díaz-Caneja CM, Ayora M, Durán-Cutilla M, Abregú-Crespo R, Ezquiaga-Bravo I, et al. Assessment of school anti-bullying interventions: a meta-analysis of randomized clinical trials. *JAMA Pediatrics*. (2020) 175:44–55. doi: 10.1001/jamapediatrics.2020.3541
10. Ikeda T, Hori D, Yu I, Takahashi T, Shiraki N, Andrea CS, et al. School ijime (Bullying) experience is a possible risk factor for current psychological distress among science city workers: a cross-sectional study in Tsukuba, Japan. *Tohoku J Exp Med*. (2020) 250:223–31. doi: 10.1620/tjem.250.223
11. Katz AJ, Hensel DJ, Zaban L, Hensley MM, Ott MA. 141 School connectedness protective against bullying in a rural community. *J Adolesc Health*. (2019) 64. doi: 10.1016/j.jadohealth.2018.10.157
12. Hansson E, Garmy P, Vilhjálmsson R, Kristjánsdóttir G. Bullying, health complaints, and self-rated health among school-aged children and adolescents. *J Int Med Res*. (2020) 48:300060519895355. doi: 10.1177/0300060519895355
13. Proulx CN, Coulter R, Egan JE, Matthews DD, Mair C. Associations of lesbian, gay, bisexual, transgender, and questioning–inclusive sex education with mental health outcomes and school-based victimization in US high school students. *J Adolesc Health*. (2019) 64:608–14. doi: 10.1016/j.jadohealth.2018.11.012
14. Duarte-Rodrigues L, Ramos-Jorge ML, Alves-Duarte AC, Fonseca-Silva T, Flores-Mir C, Marques LS. Oral disorders associated with the experience of verbal bullying among Brazilian school-aged children a case-control study. *J Am Dental Assoc* 1939. (2020) 151:399–406. doi: 10.1016/j.adaj.2020.02.001
15. Juvonen J, Schacter HL. When low rates of bullying increase risks for those who are bullied: the safe school paradox. *JAMA Pediatrics*. (2020) 174:317–8. doi: 10.1001/jamapediatrics.2019.5888
16. Harris HS, Kim, CJ. Bullying victimization, school environment, and suicide ideation and plan: focusing on youth in low-and middle-income countries. *J Adolesc Health*. (2020) 66:115–22. doi: 10.1016/j.jadohealth.2019.07.006
17. Ryzin M, Roseth CJ. Effects of cooperative learning on peer relations, empathy, and bullying in middle school. *Aggress Behav*. (2019) 45.doi: 10.1002/ab.21858
18. Kriech AC, Devon MS. 269 Bullying, weapon carrying and mental health outcomes among U.S. high school students - results from a nationally representative survey. *J Adolesc Health*. (2019) 64.doi: 10.1016/j.jadohealth.2018.10.286
19. Kitchen J. Theatre and drama education and populism: the ensemble 'family' as a space for dialogic empathy and civic care. *Br Educ Res J*. (2020) 47.doi: 10.1002/berj.3668
20. Tok HH, Cerit B. The effect of creative drama education on first-year undergraduate nursing student attitudes toward caring for dying patients. *Nurse Educ Today*. (2021) 97:104696. doi: 10.1016/j.nedt.2020.104696
21. Baselga SV, Garrido OM, Burón HG. Drama-based activities for STEM education: encouraging scientific aspirations and debunking stereotypes in secondary school students in Spain and the UK. *Res Sci Educ*. (2022) 52. doi: 10.1007/s11165-020-09939-5
22. Riley AH, Sood S, Ani MS. Narrative persuasion and social norms in entertainment-education: results from a radio drama in Mozambique. *Health Commun*. (2019) 35:1023–32. doi: 10.1080/10410236.2019.1606137
23. Cofré H, Núñez P, Santibáñez D, Pavez JM, Valencia M, Vergara C. A critical review of students' and teachers' understandings of nature of science. *Sci Educ*. (2019) 28:205–48. doi: 10.1007/s11191-019-00051-3
24. Utuk Z, Gorp KV. Putting process drama in practice with TBLT principles: TESOL in action. *TESOL Quarterly*. (2021) 55:643–54. doi: 10.1002/tesq.3003
25. Baselga SV, Garrido OM, Burón HG. Correction to: drama-based activities for STEM education: encouraging scientific aspirations and debunking

stereotypes in secondary school students in Spain and the UK. *Res Sci Educ*. (2022) 52:749. doi: 10.1007/s11165-020-09968-0
26. Ljunggren C, Carlson E, Isma GE. Drama with a focus on professional communication – A phenomenographic study. *Nurse Educ Pract.* (2021) 52:103022. doi: 10.1016/j.nepr.2021.103022
27. Enol FB, Yüksel A, Akyol T. Evaluation of perceptions of preschool education teacher candidates towards drama. *Turkish Stud*. (2019) 14:873–91.
28. Musicka-Williams A. "We Copy to Join in, to Not Be Lonely": adolescents in special education reflect on using dramatic imitation in group dramatherapy to enhance relational connection and belonging. *Front Psychol*. (2020) 11:588650. doi: 10.3389/fpsyg.2020.588650
29. Moreno-Guerrero AJ, Parra-González ME, López-Belmonte J, Robles AS. Science mapping analysis of "cultural" in web of science (1908–2019). *Qual Quant*. (2022) 56:1–19. doi: 10.1007/s11135-021-01131-9
30. Pauksztat B, Denise S. Targets' social relationships as antecedents and consequences of workplace bullying: a social network perspective. *Front Psychol*. (2019) 10:3077. doi: 10.3389/fpsyg.2019.03077

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fpubh.2022.916329

The Relationship Between Adjustment and Mental Health of Chinese Freshmen: The Mediating Effect of Security and the Moderating Effect of Gender

*Liu Cao**

School of Physical Education, Xuzhou University of Technology, Xuzhou, China

During the COVID-19 pandemic, many universities are confronted with campus lockdown or even school closures to reduce the risk of infection. However, these measures pose a threat to the mental health of adolescents. In particular, freshmen who have just entered the university campus may suffer from more serious mental health risks. In this study, 1,818 freshmen were analyzed by using the Chinese College Student Adjustment Scale (CCSAS), Sense of Security Questionnaire (SQ), and Symptom Check List 90 (SCL-90) of the qualitative method. The results showed that adjustment had an impact on mental health. Firstly, there was a significant negative association between adjustment and mental health. Secondly, adjustment had a significant predictive effect on mental health. Finally, a sense of security and gender affected the relationship between adjustment and mental health. Collectively, adjustment, sense of security, and gender exerted impacts on the mental health of freshmen, suggesting that we should create a warm and harmonious campus environment for students and conduct targeted education for male and female students.

Keywords: adjustment, mental health, security, moderating effect, mediating effect

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Srinivasa Rao Namburi,
Bapatla Engineering College, India
Rmahammad Shafi,
Mizan Tepi University, Ethiopia

***Correspondence:**
Liu Cao
274220929@qq.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 April 2022*
Accepted: *02 May 2022*
Published: *23 May 2022*

Citation:
Cao L (2022) The Relationship Between Adjustment and Mental Health of Chinese Freshmen: The Mediating Effect of Security and the Moderating Effect of Gender. Front. Public Health 10:916329. doi: 10.3389/fpubh.2022.916329

INTRODUCTION

Data from recent years indicate that mental health issues have constituted significant burdens affecting a considerable number of people worldwide. Particularly, in the context of the novel coronavirus disease 2019 (COVID-19) pandemic, the mental health and wellbeing of people are experiencing critical challenges, leading to a global mental health crisis (1). A variety of COVID-19 pandemic-triggered mental health problems has been reported in different groups in society, such as stress, anxiety, depression, and sleep disorders (2–4). What's more, negative emotions and related problems in daily life, work, or study continue to rise with the spread of the epidemic. To curb the spread of the virus, many countries have to practice different social distancing measures, such as a lockdown of the campus and a shift to distance or online learning (5). Sifat et al. (6) have noted that adolescents are more susceptible to mental health issues during the lockdown. University freshmen are a special population enduring a critical period of exploring the campus and socializing with peers. However, the school closure brought by the COVID-19 pandemic forces them to adapt to online learning and exams and reduce interactions with peers (7), which is unfavorable to the mental health of adolescents and psychosocial development (8). Therefore, the research on the

relationship between adjustment and mental health has received increasing academic attention. The present study aims to investigate the relationship between adjustment and mental health among Chinese university freshmen, hoping to provide a basis for the school and society to give accurate psychological aid to university students.

Adjustment is primarily the response to psychosocial stressors or multiple stressors, which is accepted as an important response indicator of mental health. According to the diagnostic criteria provided in the Diagnostic and Statistical Manual of Mental Disorders-5th edition (DSM-5), individuals with adjustment disorders are accompanied by significant emotional and behavioral symptoms (9). Evidently, there is a high correlation between adjustment and mental health. Accumulating studies have elucidated the significant negative correlation between mental health and adjustment in college students (10–12). Such negative correlation has also been validated in other groups. For instance, the study by Melero et al. (13) has suggested that adults who are adopted during childhood tend to develop more adjustment problems and present a higher incidence of depression, anxiety, and personality and behavioral disorders than the general population. A study of young adults who have experienced stress events demonstrates that higher levels of mental health are correlated with fewer symptoms of adjustment disorders (14). A literature review has revealed a negative relationship between adjustment and mental health, but there are few studies on freshmen. Herein, this study proposes hypothesis 1: adjustment exerts a significant effect on the mental health of freshmen.

A sense of security is considered to be an important influencing factor in regulating the relationship between adjustment and mental health (15). A sense of security is a sense of control over life, which reflects an individual's ability to deal with life problems and the belief that an individual is liked and accepted by himself and others (16). Maslow believes that people with a low sense of security tend to remain alert, feel isolated, and often exhibit tension and introversion (17). Emotional Security Theory states that maintaining and preserving a sense of security is the most basic task of an individual (18). When a negative interpersonal environment (e.g., interpersonal blocking) undermines personal security, individuals take various measures to defuse or avoid these threats to maintain their security (18, 19). Accordingly, Jia et al. (20) have noted that a sense of security is a key mediator between adjustment (e.g., interpersonal relationships) and problem behaviors. Geng et al. (21) have also found that a sense of security is a mediator between interpersonal sensitivity (e.g., shyness) and depressive symptoms. Moreover, the mediating role of security between peer relationships and internet addiction has been documented (20). The generation of a sense of security is a complex process, in which the environment is a pivotal influencing factor that has an impact on the individual's physiology and psychology. Among students, those with a higher sense of security and belonging have a higher level of mental health (22). Therefore, based on the Emotional Security Theory, the study proposes hypothesis 2: a sense of security plays a mediating role in the relationship between adjustment and mental health.

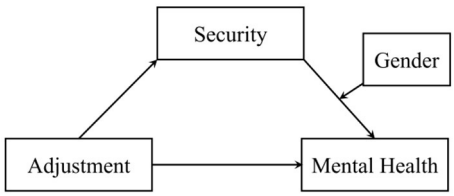


FIGURE 1 | Relationship between adjustment and mental health: the effect of security and gender.

Age and gender are significant psychological and physiological variables (23). During the COVID-19 pandemic, adolescents suffer more social isolation and family stress and experience more depression and anxiety, and furthermore, the severity of mental health problems in women is much higher than that in men (4). Accordingly, this study proposes hypothesis 3: gender moderates the relationship between a sense of security and mental health. Based on the hypotheses, this study proposes a conceptual model map, as shown in **Figure 1**.

MATERIALS AND METHODS

Participants

The freshmen from six universities in Jiangsu Province were investigated by random sampling. This study was approved by the college ethics committee. The subjects were informed of the purpose of the survey and informed consent was obtained before the survey. A total of 2,000 questionnaires were distributed and 1,873 were received, with a questionnaire response rate of 93.7%; the final number of valid subjects was 1,818, with an effective rate of 97.1%. Among the valid subjects, 816 (44.9%) were male and 1,002 (55.1%) were female, ranging from 17 to 24 years old, with an average age of 19.47 ± 1.18 years old.

Instruments

Chinese College Student Adjustment Scale (CCSAS), Sense of Security Questionnaire (SQ), and Symptom Check List 90 (SCL-90) were adopted.

(a) Chinese College Student Adjustment Scale (CCSAS). Fang et al. (24) prepared the scale with 60 items, including 7 dimensions of interpersonal adjustment, academic adjustment, campus life adjustment, career choice adjustment, emotional adjustment, self-adjustment, and satisfaction. A 5-point scale was used, ranging from 1 point (disagree) to 5 points (agree), but some entries were reverse-scored. A higher score was indicative of a higher level of adjustment. This scale had good reliability and validity, with the Cronbach's alpha ranging from 0.67 to 0.82 for the 7 dimensions and 0.93 for the total scale.

(b) Sense of Security Questionnaire (SQ). Cong and An (25) formulated the scale with 16 items, which were divided into 2 factors: interpersonal security and certainty in control. The scale was scored on a 5-point scale from 1 point (very consistent) to 5 points (very inconsistent), with a higher

TABLE 1 | Comparison of the differences in SCL-90, CCSAS, and SQ among freshmen of different genders (n = 1,818).

Variables	Female	Male	t
SCL-90	1.70 ± 0.43	1.57 ± 0.40	6.67***
CCSAS	3.31 ± 0.55	3.39 ± 0.56	−3.17**
SQ	3.55 ± 0.67	3.75 ± 0.68	−6.33***

**p<0.01; ***p<0.001.

score indicating a higher sense of security. The Cronbach's alpha was 0.75 for interpersonal security, 0.72 for certainty in control, and 0.80 for the total scale.

(c) Symptom Check List 90 (SCL-90). The scale was prepared by Derogatis and revised by Liu and Zhang (26). The scale contained a total of 90 items, including 9 factors such as somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobia anxiety, paranoid ideation, and psychoticism, and 1 other factor, and these 10 factors constituted the total mental health score. A 5-point scale was used, ranging from 1 point (none) to 5 points (severe). Higher factor scores and total scores were associated with more prominent psychological problems. The Cronbach's alpha of the 10 factors for secondary school students ranged from 0.70 to 0.88.

Procedures

The test was conducted by team members who had received systematically training and obtained a qualification in measurement before the test. The test was conducted as a group of a class, and the time limit was about 30 min. All questionnaires were distributed and recovered on the spot. The recovered questionnaires were checked by a team member and then a database was created. The data were analyzed and processed using SPSS27.0 and the Macro Program Process developed by Hayes. The gender differences in adjustment, sense of security, and mental health were analyzed using the independent sample t-test, and the correlation between adjustment, sense of security, and mental health was analyzed using Pearson product-moment correlation coefficient. The mediating effect of security and the moderating effect of gender were analyzed using the Macro Program Process developed by Hayes. The bootstrap method was used, with a sample size of 5,000 and a confidence interval of 95%.

RESULTS

Difference Comparison

As shown in **Table 1**, there were significant differences in the scores of SCL-90, CCSAS, and SQ scales between male and female students. Specifically, female students got higher scores than male students in SCL-90, and male students got higher scores than female students in CCSAS and SQ. According to the difference comparison results and previous analysis, it could be concluded that gender played a moderating role in the relationship between SQ and SCL-90.

TABLE 2 | Descriptive statistics and correlation analysis of the variables.

	M	SD	SCL-90	CCSAS	SQ	Gender
SCL-90	1.64	0.42	1			
CCSAS	3.35	0.56	−0.49**	1		
SQ	3.64	0.68	−0.60**	0.67**	1	
Gender	–	–	0.07**	0.15**	−0.16**	1

**p<0.01.

TABLE 3 | Relationship between CCSAS and SCL-90 in freshmen: the model of SQ and gender.

Predictor variable	Outcome variable	R	R^2	F	β	t	LLCI	ULCI
Equation 1								
CCSAS	SQ	0.67	0.45	1,455.58***	0.67	38.15***	0.63	0.70
Equation 2								
CCSAS	SCL-90	0.62	0.38	281.04***	−0.17	−6.13***	−0.22	−0.12
SQ					−0.52	−16.99***	−0.58	−0.46
Gender					−0.15	−3.92**	−0.22	−0.07
SQ × gender					0.09	2.41*	0.02	0.16

*p<0.05; **p<0.01; ***p<0.001.

Descriptive Statistics and Correlation Analysis

The correlation analysis revealed (**Table 2**) that CCSAS, SQ, and SCL-90 were significantly correlated with gender. CCSAS, SQ, and SCL-90 showed a significant negative correlation, while CCSAS and SQ showed a significant positive correlation.

Moderating and Mediating Effects Analyses

According to the test method of the structural equation model and the results of correlation analysis, the conditions for testing moderated mediation model were satisfied, so the mediating role of SQ and the moderating role of gender were examined in the relationship between CCSAS and SCL-90. CCSAS, SQ, and SCL-90 were standardized before the analysis, with female = 0 and male = 1, and then, the effect test was performed.

The analysis found (**Table 3**) that CCSAS, SQ, and gender had a significant main effect on SCL-90 [effect = −0.17, $p < 0.001$, 95% CI (−0.22 to −0.12); effect = −0.52, $p < 0.001$, 95% CI (−0.58 to −0.46); effect = −0.15, $p < 0.01$, 95% CI (−0.22 to −0.07)]. The main effect of CCSAS on SQ was also significant [effect = 0.67, $p < 0.001$, 95% CI (0.63–0.70)]. SQ interacting with gender showed a significant predictive effect on SCL-90 [effect = 0.09, $p < 0.05$, 95% CI (0.02–0.16)]. The main effect and interaction effect analyses indicated that the mediating role of SQ and the moderating role of gender could be further analyzed.

The analysis revealed the moderating effect of gender in the relationship between SQ and SCL-90, wherein in the male group, the effect value of SQ on SCL-90 was significant [effect = −0.52, $P < 0.001$, 95% CI (−0.58 to −0.46)] and in the female group, the effect value of SQ on SCL-90 was also significant [effect = −0.43, $P < 0.001$, 95% CI (−0.49 to −0.37)]. The slope analysis showed

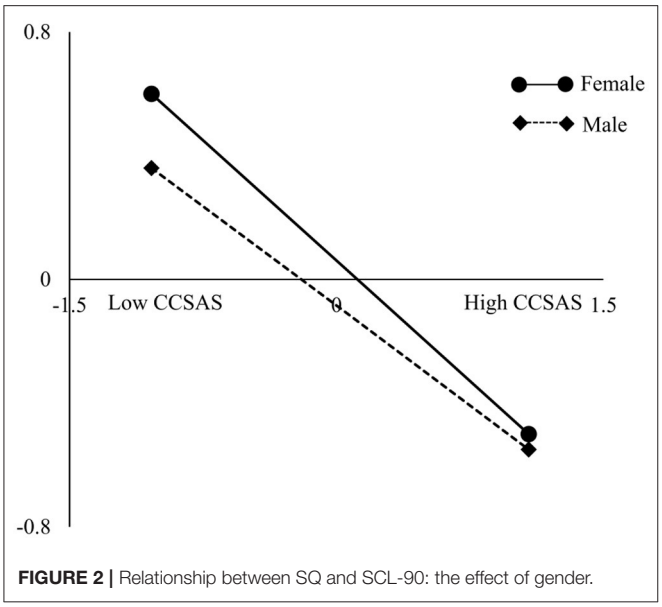


FIGURE 2 | Relationship between SQ and SCL-90: the effect of gender.

TABLE 4 | Relationship between CCSAS and SCL-90 of freshmen: the effect of SQ and gender.

	Female		Male	
	Effect	**Ratio of indirect to total effect**	**Effect**	**Ratio of indirect to total effect**
Indirect effect	−0.35	67.3%	−0.29	63%
Direct effect	−0.17	32.7%	−0.17	37%
Total effect	−0.52	–	−0.46	–

(**Figure 2**) that the level of SCL-90 was elevated with the increase of SQ, but the change in the female group was more obvious than that in the male group.

Further mediating effect analysis (**Table 4**) demonstrated that the indirect effect of CCSAS→ SQ→ SCL-90 was −0.35 for the female group, accounting for 67.3% of the total effect and the indirect effect of CCSAS→ SQ→ SCL-90 was −0.29 for the male group, accounting for 63% of the total effect. It was indicated that a sense of security was a vital mediating variable in the relationship between adjustment and mental health.

DISCUSSION

The study revealed the mechanism of adjustment in mental health. Security played a mediating role in the relationship between adjustment and mental health and such mediating effect was moderated by gender. The main contribution of this study is to analyze the relationship between adjustment and mental health, especially the role of security, and also provide further evidence for understanding the role of gender in the above relationships.

It was found that adjustment was positively correlated with a sense of security, and security was negatively correlated with mental health. The sense of security refers to an individual's sense of anticipation of possible physical and psychological risks and a sense of power/powerlessness in difficulty, mainly in the form of interpersonal security and certainty in control. Adjustment, on the other hand, is a reaction or state in the face of a stressor. When facing a new environment, individuals will make predictions and find ways to make active adjustments. If individuals can gain a sense of security during adjustment, they will continue to promote adjustment and maintain healthy psychology; if not, they will have difficulties in adjustment, leading to mental health problems. Therefore, security is a major variable linking adjustment and mental health. Consistently, Geng et al. (21) have validated Emotional Security Theory and Davies et al. (19) have stated that maladjustment is related to interpersonal stress and it is the personal security crisis caused by maladjustment and stress that leads to mental health problems. Especially university freshmen are facing interpersonal and academic adjustments, but the blockade brought by the COVID-19 pandemic triggers interpersonal and academic stress, which can lead to mental health problems such as depression and anxiety. In this process, the role of a sense of security becomes prominent. If the school gives students a sense of belonging and security and guides them to adjust positively, the emergence of mental health symptoms will be retarded (22).

More importantly, this study identified the role of gender in the relationship between mental health, adjustment, and security. Gender moderated the relationship between security and mental health, thus exerting an important influence on the relationship between the three, with greater changes in the female group in comparison to the male group. A survey has shown that the prevalence of any mood disorder in Chinese women is higher than that in men (27). Lu et al. (28) has also observed a higher prevalence of depressive disorders in Chinese women than men. Similarly, this study explored the role of gender in the relationship between CCSAS, SQ, and SCL-90 and also found greater changes in females. In the context of Chinese culture, the female is usually defined as tenderness, kindness, and introversion, among which sensitivity is a major trait. Male, on the other hand, are usually characterized as masculinity, self-confidence, and a sense of responsibility, among which perseverance is a major trait. The different male and female traits have different effects on the adjustment, sense of security, and mental health of students. In addition, in Chinese society and culture, men have more social and competitive opportunities than women, which also leads to gender differences. It is evident that gender continues to be an important factor in life change in China.

CONCLUSION

This study explores the relationship between adjustment and mental health and determines the multiple effects of security and gender. The relationship between adjustment and mental health can be mediated through a sense of security, and gender moderates the relationship between security and mental health. The findings of this study provide a direction for the current mental health education.

Based on the above findings, we call on universities and relevant responsible departments to make full use of social networking platforms to guide freshmen to communicate and also create a healthy and positive social networking environment to cope with the obstacles brought by school closures. Teaching activities and social communication activities should be actively carried out within the limited space to enable freshmen to integrate into the new environment as soon as possible. On the premise of ensuring safety, universities should actively organize various campus activities to enrich students' after-school life and help them cope with the boring and isolated campus life. In addition, our results suggest that future school mental health education should be committed to creating an environment with a sense of security and belonging, and provide different educational activities for male and female students.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of School of Physical Education, Xuzhou Institute of Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LC: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript.

FUNDING

This work was supported in part by the Xuzhou Science and Technology Plan Project under Grant KC21303, Jiangsu Industry University Research Cooperation Project under Grant BY2021159, and the sixth 333 project of Jiangsu Province.

REFERENCES

1. Abbott A. COVID's mental-health toll: how scientists are tracking a surge in depression. *Nature.* (2021) 590:194–5. doi: 10.1038/d41586-021-00175-z
2. Ahmed F, Sifat RI. The impact of the COVID-19 pandemic on the mental health of the Rickshaw-Puller in Bangladesh. *J Loss Trauma.* (2021) 26:782–9. doi: 10.1080/15325024.2020.1852823
3. Beal JA. Impact of the covid-19 pandemic on mental health of children, adolescents, and young adults. *MCN Am J Matern Child Nurs.* (2021) 46:237. doi: 10.1097/NMC.0000000000000726
4. Meade J. Mental health effects of the covid-19 pandemic on children and adolescents: a review of the current research. *Pediatr Clin North Am.* (2021) 68:945–59. doi: 10.1016/j.pcl.2021.05.003
5. Rodriguez RJE, Oviedo-Trespalacios O, Cortes-Ramirez J. A brief-review of the risk factors for covid-19 severity. *Revista de Saúde Pública.* (2020) 54:60. doi: 10.11606/s1518-8787.2020054002481
6. Sifat RI, Ruponty MM, Shuvo KR, Chowdhury M, Suha SM. Impact of COVID-19 pandemic on the mental health of school-going adolescents: insights from Dhaka City, Bangladesh. *Heliyon.* (2022) 8:e09223. doi: 10.1016/j.heliyon.2022.e09223
7. UNICEFIRC. *Promising Practices for Equitable Remote Learning. Emerging Lessons From COVID-19 Education Responses in 127 Countries*. (2020). Available online at: https://www.unicef-irc.org/publications/1090-promisingpractices-for-equitable-remote-learning-emerging-lessonsfrom-covid.html (accessed 10, 2020).
8. Lewis SJ, Munro A, Smith GD, Pollock AM. Closing schools is not evidence based and harms children. *BMJ.* (2021) 2:372. doi: 10.1136/bmj.n521
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders(DSM-5)* Washington, DC (2013).
10. Dong Y, Dang L, Li S, Yang X. Effects of facets of mindfulness on college adjustment among first-year chinese college students: the mediating role of resilience. *Psychol Res Behav Manage.* (2021) 14:1101–9. doi: 10.2147/PRBM.S319145
11. Zhang G, Yang X, Tu X, Ding N, Lau JTF. Prospective relationships between mobile phone dependence and mental health status among Chinese undergraduate students with college adjustment as a mediator. *J Affect Disord.* (2020) 260:498–505. doi: 10.1016/j.jad.2019.09.047
12. Zhang G, Tu X, Ding N, Lau JTF, Wang P, Yang X. Prospective relationships between college adjustment, self-esteem, and mental health status among Chinese undergraduates. *J Am College Health.* (2021) 9:1–7. doi: 10.1080/07448481.2021.1909036
13. Melero S, Sánchez-Sandoval Y. Mental health and psychological adjustment in adults who were adopted during their childhood: a systematic review. *Children Youth Serv Rev.* (2017) 77:188–96. doi: 10.1016/j.childyouth.2017.05.006
14. Truskauskait-Kuneviien I, Kazlauskas E, Ostreikaite-Jurevice R, Brailovskaia J, Margraf J. Positive mental health and adjustment following life-stressors among young adults. *Curr Psychol.* (2020) 3:1–6. doi: 10.1007/s12144-020-00714-3
15. Jia J, Li D, Li X, Zhou Y, Wang Y, Sun W. Psychological security and deviant peer affiliation as mediators between teacher-student relationship and adolescent Internet addiction. *Comput Human Behav.* (2017) 73:345–52. doi: 10.1016/j.chb.2017.03.063
16. Jourard SM, Remy RM. Perceived parental attitudes, the self, and security. *J Consult Psychol.* (1955) 19:364e366. doi: 10.1037/h0049318
17. Maslow AH. The dynamics of psychological security-insecurity. *J Personality.* (1942) 10:331e344. doi: 10.1111/j.1467-6494.1942.tb01911.x
18. Cummings EM, Miller-Graff LE. Emotional security theory: an emerging theoretical model for youths' psychological and physiological responses across multiple developmental contexts. *Curr Direct Psychol Sci.* (2015) 24:208e213. doi: 10.1177/0963721414561510
19. Davies PT, Martin MJ, Sturge-Apple ML. Emotional security theory and developmental psychopathology. In Cicchetti D, editor, *Developmental Psychopathology: Vol. 1. Theory and Methods*. (2016) (New York, NY: Wiley). p. 199–264.
20. Jia J, Li D, Li X, Zhou Y, Wang Y, Sun W, et al. Peer victimization and adolescent Internet addiction: the mediating role of psychological security and the moderating role of teacher-student relationships. *Comput Human Behav.* (2018) 85:116–24. doi: 10.1016/j.chb.2018.03.042
21. Geng J, Li L, Han LC, Gao F. Shyness and depressive symptoms: a multiple mediation model involving core self-evaluations and sense of security. *J Affect Disord.* (2021) 286:19–26. doi: 10.1016/j.jad.2021.01.035
22. Leurent B, Dodd M, Allen E, Viner R, Scott S, Bonell C. Is positive school climate associated with better adolescent mental health? Longitudinal study of young people in England. *SSM Mental Health.* (2021) 1:100033. doi: 10.1016/j.ssmmh.2021.100033

23. McGue M, Bouchard TJ. Adjustment of twin data for the effects of age and sex. *Behav Genet.* (1984) 14:325–43. doi: 10.1007/BF01080045
24. Fang XY, Wo JZ, Lin XY. Development of China college student adjustment scale. *Stud. Psychol. Behav.* (2005) 3:95–101.
25. Cong Z, An L. Developing of security questionnaire and its reliability and validity. *Chinese Mental Health J.* (2004) 2:97–9.
26. Liu H, Zhang JX. Norm of symptom check list (SCL-90) in Chinese middle school students. *Chinese Mental Health J.* (2004) 2:88–90.
27. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. *Lancet Psychiatry*. (2019) 6:211–24. doi: 10.1016/S2215-0366(18)30511-X
28. Lu J, Xu X, Huang Y, Li T, Ma C, Xu G, et al. Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. *Lancet Psychiatry*. (2021) 8:981–90. doi: 10.1016/S2215-0366(21)00251-0

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fpubh.2022.909628

Deep Conviction Systems for Biomedical Applications Using Intuiting Procedures With Cross Point Approach

Hariprasath Manoharan[1], Shitharth Selvarajan[2]*, Ayman Yafoz[3], Hassan A. Alterazi[3], Mueen Uddin[4], Chin-Ling Chen[5,6,7]* and Chih-Ming Wu[8]*

[1] Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai, India, [2] Department of Computer Science & Engineering, Kebri Dehar University, Kebri Dehar, Ethiopia, [3] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, [4] School of Digital Science, University Brunei Darussalam, Bandar Seri Begawan, Brunei, [5] School of Information Engineering, Changchun Sci-Tech University, Changchun, China, [6] Department of Computer Science and Information Engineering, Chaoyang University of Technology, Taichung, Taiwan, [7] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China, [8] School of Civil Engineering and Architecture, Xiamen University of Technology, Xiamen, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Sundaravadivazhagan Balasubaramainan, University of Technology and Applied Sciences, Oman
K. Anitha Kumari, PSG College of Technology, India
G. Surya Narayana, Vardhaman College of Engineering, India

***Correspondence:**
Hariprasath Manoharan
hari13prasath@gmail.com
Shitharth Selvarajan
shitharth.it@gmail.com
Chin-Ling Chen
clc@mail.cyut.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *31 March 2022*
Accepted: *14 April 2022*
Published: *23 May 2022*

Citation:
Manoharan H, Selvarajan S, Yafoz A, Alterazi HA, Uddin M, Chen C-L and Wu C-M (2022) Deep Conviction Systems for Biomedical Applications Using Intuiting Procedures With Cross Point Approach. Front. Public Health 10:909628. doi: 10.3389/fpubh.2022.909628

The production, testing, and processing of signals without any interpretation is a crucial task with time scale periods in today's biological applications. As a result, the proposed work attempts to use a deep learning model to handle difficulties that arise during the processing stage of biomedical information. Deep Conviction Systems (DCS) are employed at the integration step for this procedure, which uses classification processes with a large number of characteristics. In addition, a novel system model for analyzing the behavior of biomedical signals has been developed, complete with an output tracking mechanism that delivers transceiver results in a low-power implementation approach. Because low-power transceivers are integrated, the cost of implementation for designated output units will be decreased. To prove the effectiveness of DCS feasibility, convergence and robustness characteristics are observed by incorporating an interface system that is processed with a deep learning toolbox. They compared test results using DCS to prove that all experimental scenarios prove to be much more effective for about 79 percent for variations with time periods.

Keywords: biomedical signals, deep learning, Fourier filters, sensors, cross points

SURVEY OF CONVENTIONAL MODELS

As many researchers have contributed their works for changing advancement that is associated with real-time effects, this part describes and contrasts numerous models that are related to developments in the biomedical field industry. Several modes of operation have been investigated even in past generations without the use of computer-aided design, and the end-stage results are insufficient for confirming the effective models. As a result, the colloquium began with a communication and signal processing toolbox (1), which uses a machine learning technique with a complicated dynamic nature and distributed optimization. However, in all case studies, a random technique is used, with a focus on sixth-generation networks, and even high sensing conditions have been made practical using these types of ideal procedures. A comprehensive review

strategy is established using several case studies, including neuron science and cardiovascular image processing techniques, where significant influence has been demonstrated under various algorithmic settings (2). However, because analytic information is required for all representation scenarios, photos are not processed using the training model. Even though 60% of the reference model is operational, the presentation of biological disease with various picture models cannot be compared under extreme situations.

Aside from the case studies stated above, a deep learning method with high feature extraction has been shown to be an effective technique for geometric scale analysis (3). The main drawback of geometric scale analysis is that dynamic state measurements do not work with baseline health data because differences are noticed at stage 2. Even feature state analysis has been investigated using weight transformations, and such state changes have been shown to operate with a variety of deep representations. However, data augmentation using end networks only works when there are numerous learning problems (4), which are not present in this type of deep learning technique. Furthermore, the physiological expression of signals is categorized using random value creation, and the values are observed, resulting in heavy noisy data (5). As signals are reconstructed as the value pushes lower down to the source point, this noisy data will have a significant impact on biomedical signals, with more physiological sources having computational consequences as signals are recreated. As morphological data is varied throughout separate time intervals, the aforementioned drop in source point can be convalesced using a soft max function. Data will be split into five separate forms for soft max functions, enhancing the capability of the abstraction process.

Furthermore, the firefly algorithm (6) can be used to handle issues and challenges linked to biological applications, where all health problems are studied in-depth, resulting in creative answers through data fusion. Many parametric measures are still displaced due to insufficient access information from other networks, necessitating realistic experimentation with data generated from a centralized perspective. After detecting minimal parametric changes in the firefly method, real-time signal generator algorithm updates are made with incoming signal validation (7). Protocols are also framed for the generation of algorithms using the layered decomposition technique, which allows signals to be partially communicated. This partial signal transfer will be accomplished using a decoding technique that is based on binary values. As a result, a unique formulation is required for converting a standard arrangement to a binary system, which can be accomplished using training procedures seen in transmission lines with epoch iterations. Despite the fact that real-time algorithms are more efficient than 70% of the time, data sets are not established in a good initialization model, thus a primate algorithm with high extraction features and parametric stabilizations is used with a longer overall testing time (8). Due to the fact that stabilizations are prepared in a dual mode of operation, relative errors are further lowered to optimum values, which on average fall below 82. Despite the intricacy of temporal representations, two different data sets will be presented, enhancing the risk of misinterpretation during extensive analysis.

Furthermore, biological signals are integrated at a massive scale utilizing gate count technology (9), and various signals are handled using system architectures. The operating frequencies of various signals are not determined due to the presence of diverse designs, resulting in a miscount of separate integration gates. As a starting point, neural networks are useful for processing crucial signals because they take advantage of a variety of compositional properties of biological points of representation (10). Furthermore, if any problems are found in a large structure, ECG and EEG data will be isolated, allowing for future orientations and change. Between 2012 and 2018, there were numerous techniques that used sensor equipment and permitted the feel and touch method during the communication process (11). Electroencephalography data are evaluated in this feeling, which is modeled as a new technique with seven distinct layers. In accordance with the aforementioned issue, the 7-layer technique is carried out in parallel units, with programming written at high rates employing processor units. Many advantages can be seen for biomedical applications with such high rates, where justification can be processed before pre-processing phases (12–14). These advantageous systems are created using wearable devices where auto-learning models are activated. This type of automated process will allocate signals that move at different frequency periods with a classification mechanism (15). However, it is much easier to separate different signals at unique frequency periods but a wearable device must be created for separating the individual monitored values rather than signals. Further biomedical signals are tested for the prediction of cardiovascular disease using conventional neural networks where a short-term memory coder is implemented in the system (16, 17). The major disadvantage is that if signals are processed, then high memory is required for storing the information that can be accessed at later stages.

Research Gap and Motivation

All the different approaches (1–14) that are discussed using different methods for processing biomedical signals which are not accurate in the field of medical diagnosis. The biomedical signals cannot be passed without the presence of proper communication units at both the transmitter and receiver sides. However few efforts (15–17) have been made by some researchers to prove the complete effectiveness of biomedical signal processing in various applications and they have achieved 63% of accuracy in considered cases. Therefore, the proposed method witnesses the abovementioned gap and provides a solution for effectively utilizing biomedical signals in medical applications. The process functions using a Deep Conviction Systems (DCS) which is an automated procedure for capturing the image of a particular individual using low power technology. This type of implementation system reduces the risk of implementation as high resistance is provided with tracking mechanisms. Also, the data gathering path using biomedical signals is faster and it is replaced with some units in the existing system. This adds an advantage as the entire system is built with low-cost networks

that provide more than 80% accuracy on the input side even if the signal strength is at the intermediate stage.

Objectives

The proposed work on biomedical signal processing is a three-stage objective process that aims to solve the following,

- To incorporate a pre-processing stage that incarcerations the input images to a small matrix type
- To maximize the distance of measurement by transmitting the biomedical signals with recorded output type
- To generate a data augmentation procedure for apprehending the information that is processed using DCS

SYSTEM MODEL: GEOMETRIC ANALYSIS OF SIGNALS

In this section, the system model that supports the generation of signals with respect to algebraic equations is framed using two types of instability models. The major reason for selecting unsteady models is that the signals will be generated using the spectral mode of transmission and it varies for several operating cases. Thus, a segment of sensors is integrated for monitoring huge variations of input signals therefore all recorded output signals will be represented using a non-specific Equation which is represented below,

$$d_a(t) = \sum_{i=1}^{n} s_t\,(i) + n_i(t) \tag{1}$$

Where,

$d_a(t)$ indicates the acquired data that changes with time periods

$s_t\,(i)$ and $n_i\,(t)$ represents true stable and noisy signals respectively

From Equation (1) the noisy signals will have Fourier filters as discrete noise from individual signals is divided with time period representations. Further filters are applied in the case where different time scales are observed for periodic scales as biomedical signals are supplied with a data attainment approach. Thus the time scales can be varied using Equation (2) as follows,

$$t_i = \sum_{i=1}^{n} d_a\left(t - \frac{s_t}{2}\right)\left(t + \frac{s_t}{2}\right) \tag{2}$$

Where,

$t - \frac{s_t}{2}$ and $t + \frac{s_t}{2}$ represents positive and negative time samples

Equation (2) is represented in the form of significant variant properties thus the concession matrix can be used for normalizing all the connected signals. This can be represented in mathematical form as follows,

$$\rho_i = \frac{1}{2} \sum_{i=1}^{n} w_i\,|c_{in}| \tag{3}$$

Where,

w_i denotes weight factor of applied bio-signals

c_{in} indicates the number of connected networks in the signaling process

The connected networks in the signal processing technique will have upper and lower boundaries which are represented using shifting boundaries using Equation (4) as follows,

$$c_{in} = \sum_{i=1}^{n} \frac{u_{in}\,(t) + l_{in}(t)}{Total\ period} \tag{4}$$

Where,

$u_{in}\,(t)$ and $l_{in}\,(t)$ represents the upper and lower limits

Equation (4) is also characterized as mean values of signals and appropriate integer values cannot be obtained in any case. Therefore, the quantization of integers at each step period can be represented using Equation (5) as follows,

$$q_i\,(t) = \sum_{i=1}^{n} \left[\frac{min(in)}{max(in)}\right] * (S_i - 1) \tag{5}$$

Where,

$\min(in)$ and $\max(in)$ denotes minimum and maximum quantization values of signals

S_i represents the data size of signals

In Equation (5) data size of signals is considered using an array of the matrix as 16*16 as the convolution of signals can be carried within the corresponding network frame. Subsequently, the defined array size can provide errors and it can be solved using the probability density function of signals as represented in Equation (6) as follows,

$$\sigma_i = \sum_{i=1}^{n} \log P_t(i)(1 - \log P_o(i)) \tag{6}$$

Where,

$P_t(i)$ and $P_o(i)$ represent true probability and original values of signal positions

The geometric analysis of biomedical signals is used for solving both minimization and maximization problems that can be represented using the objective function as given in Equation (7).

$$Obj_i = min \sum_{i=1}^{n} C_{in}, t_i\ and\ max \sum_{i=1}^{n} dist_i \tag{7}$$

Equation (7) indicates that two minimizations and one maximization problem will be solved using the proposed DCS algorithm for processing the biomedical signals at high strength to monitor all necessary parameters.

OPTIMIZATION ALGORITHM

The major advantage of DCS as compared with other non-fusion algorithms is that it is not highly robust to noise and the learning rate of DCS is much higher as compared with other algorithms. Since there is a necessity to learn and update different characteristics of biomedical signals as it varies for every time period a DCS will be much support to correct the raw data behavior (18). The process involved in DCS saves the learning time as multi-tasking is performed with automatic feature learning behavior as the structure of DCS is designed in an anatomical way (19). Deep learning methods are used

for biomedical signal processing in this part because many significant features may be extracted within reasonable bounds. Even with the use of a deep learning algorithm, a wide range of complex data can be extracted using various learning patterns. As a result, one sort of deep learning algorithm known as a DCS is used since numerous learning parameters can be changed as time passes. In DCS, one visible and hidden vector is chosen at the outset, and the energy parameter is represented using a probability distribution as shown in Equation (8).

$$P\left(o_i,s_i\right) = \sum_{i=1}^{n} \frac{e^{-o_i,s_i}}{\omega_i} \tag{8}$$

Where,

o_i and s_i represents the observable and secret bound values

ω_i denotes the energy representation matrix

To make use of the energy constraint, the probability values must remain at 1. As a result, it is used as a persistent significance in observable bounds, with coordinates that can be changed in the order of j. Equation (9) can be used to represent j^{th} order variation in expectation form as follows:

$$E\left(i,n\right) = \sum_{i=1}^{n} \frac{j_i}{z_i}(1-j_i) \tag{9}$$

Where.

j_i indicates a sub-variance matrix with the change of order

In case the order in Equation (9) is recurrent then to identify the similarity index a mathematical model is framed using Equation (10) as follows,

$$j_i = \sum_{i=1}^{n} \frac{(1-\partial)^2}{\partial} log(e_{in}) \tag{10}$$

Where,

∂ denotes similarity index of the matrix

e_{in} indicates an error in the processing of similar values

Using an unvarying distributed matrix, Equation (10) shows a logarithmic control where all comparable values can be ignored. In addition, the distance between biomedical signal processing can be determined using Equation (11) as follows:

$$distance_i = \sum_{i=1}^{n} d\left(a,b\right) + d\left(x,y\right) \tag{11}$$

Where,

a, b, and x, y denote the four points of inter-section between biomedical signals

The abovementioned signals can be activated with a learning rate using maximized function values which can be defined using Equation (12).

$$r_f\left(i\right) = \begin{cases} 0 \; if \; \max(0,1) \\ 1 \; if \; \min(0,1) \end{cases} \tag{12}$$

Where,

r_f indicates the Relu function with maximization and minimization values

The Relu function in Equation (12) will be varied with binary values 0 and 1 using the tanh activation task as follows,

$$\tanh\left(i\right) = \sum_{i=1}^{n} \frac{e^i - e^n}{e^i + e^{-n}} \tag{13}$$

Where,

e^i and e^n represents tasks that are performed before and after pre-processing stages

In the presence of a distinct learning rate adjustment parameters must be defined with updating of parametric values where iteration variables are used and represented using Equation (14) as follows,

$$R_a\left(i\right) = \sum_{i=1}^{n} \frac{R_a\left(\max\right) - R_a(\min)}{number\ of\ iterations} \tag{14}$$

Where,

R_a indicates fine-tuning of parameters

After certain parametric tuning, the loop values of DCS will be produced as output with a description of the learning rate. The flow chart of DCS for biomedical signals is deliberated in **Figure 1**.

RESULTS AND DISCUSSIONS

In this section, the results of the system model presented in section 2 are simulated using an experimental setup consisting of Fourier filters and geometric signal processing. The experimental setting is constructed in such a way that the total number of physical quantities measured by a sensing device is minimized. Because polluted noise in biomedical signals is decreased, the sensors play an important role in the monitoring process. Both front and rear end sensors are used in the suggested method, and the compatibility of numerous signals is examined at the same time. A deep learning toolbox in MATLAB is used to convert the hardware setup to real-time simulation, with the major data being collected using a reference data set stored on the local disc. The following scenarios also show cross-validation of biomedical signal processing:

Scenario 1: Reconnaissance of recorded data
Scenario 2: Variation of time scale periods
Scenario 3: Configuration distance
Scenario 4: Task of activation
Scenario 5: Cost of implementation

All of the aforementioned situations are carried out with signals that are recorded in a hardware setup and compared to reference signals in computer-aided design. During this comparison, it is decided to use Fourier filters to separate the signals in a regular manner, hence decreasing noise. A full description of all scenarios will give you a good understanding of the seven-layer process.

Scenario 1

In this situation, the recorded data at output units is measured using non-specific time-series samples. Different dynamic ranges are identified without any labels of numerous

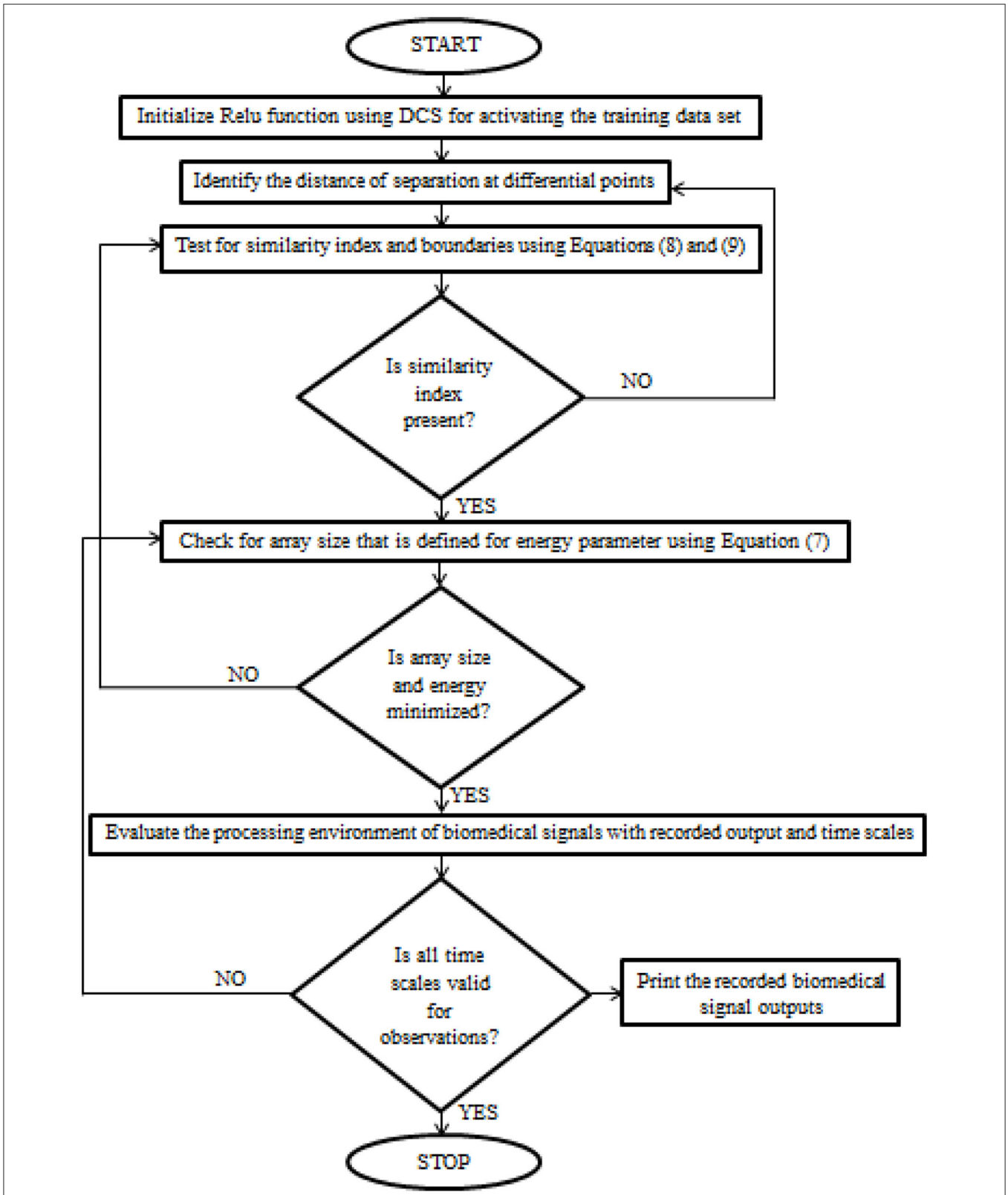


FIGURE 1 | The implementation procedure of Deep Conviction Systems (DCS) for realizing biomedical signals.

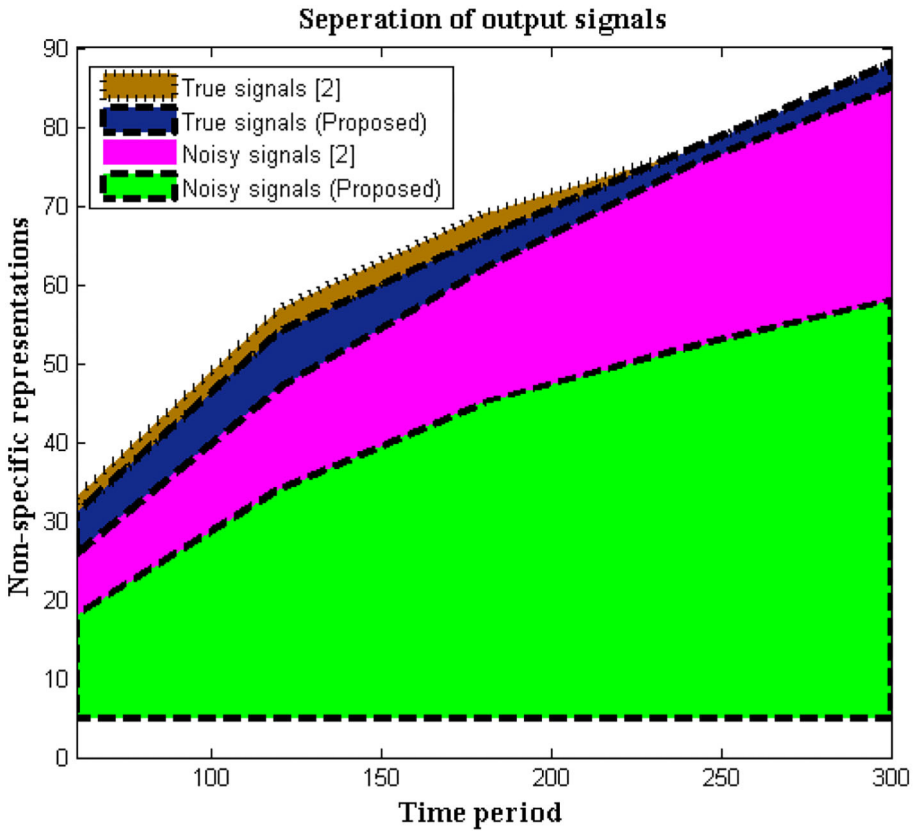


FIGURE 2 | Representation of recorded data.

signals, hence the output units are measured using a 16-bit sample period. Signals are delivered to the receiver without interruption because no tag action is done during the signaling interval. However, in biomedical signal processing, it is necessary to utilize tags to distinguish between real and noisy signals, as shown in Equation (1), and this may be viewed as a unique mechanism used in the execution of the suggested method. This type of separation protects the signal until it reaches the endpoint and allows users to choose channels based on their priority. If any caveat signals are recognized, time periods will be represented at a higher stage, resulting in a faster signal propagation speed. As limited appropriation is used, the above-mentioned technique is verified every 5 s. **Figure 2** depicts the simulation results of recorded output unit data.

Figure 2, **Table 1** show that the initial stage of representations can be achieved using these periodic representations over time durations ranging from 60 to 300 s. The number of true and noisy signals is limited throughout these periods, and the effective signal epoch boundary limits should be between 40 and 60. In addition, a comparison was done with conventional approaches using different system models, and it was discovered that the average number of real signals is formed (2), but true signal representation is determined to be at high levels in the deep learning process. In the projected technique, rather than random changes, the amount of noisy signals is reduced at the conclusion of time periods (2).

Scenario 2

It is necessary to execute the unpredictability of time periods in addition to the separation of biological signals, which is presented in this scenario. The evolution of biological signal representation is accomplished through a stationary process in which modest changes in time periods cannot be tracked over a short period of time. However, because the suggested method produces non-stationary signals, it is possible to track and adjust the signal properties even during the automatic recording step. This benefit can be used during the pre-processing step because the collected images are clear before they reach the channel state. Furthermore, this creates a lively trade-off between frequency signals, which is visible at peak values. Due to the employment of Fourier filters, a transformation strategy using sequential resolution is used, with a large window size being recommended. Furthermore, utilizing a single matrix form, the time resolution of various signals will be exploited to separate positive and negative samples. Both positive and negative signals will be normalized with a combination factor after the pre-processing stage, as shown in **Figure 3**.

The separation of data from time scales that twitch from −2 to +3 when the transformation technique is adopted is depicted in **Figure 3**, **Table 2**. Additionally, the same amount of positive

TABLE 1 | Separation of stable and noisy signals.

Time period	Number of stable signals (2)	Number of stable signals (Proposed)	Number of noisy signals (2)	Number of noisy signals (Proposed)
60	8	18	5	2
120	13	34	7	3
180	17	45	4	3
240	23	52	2	0
300	27	58	3	0

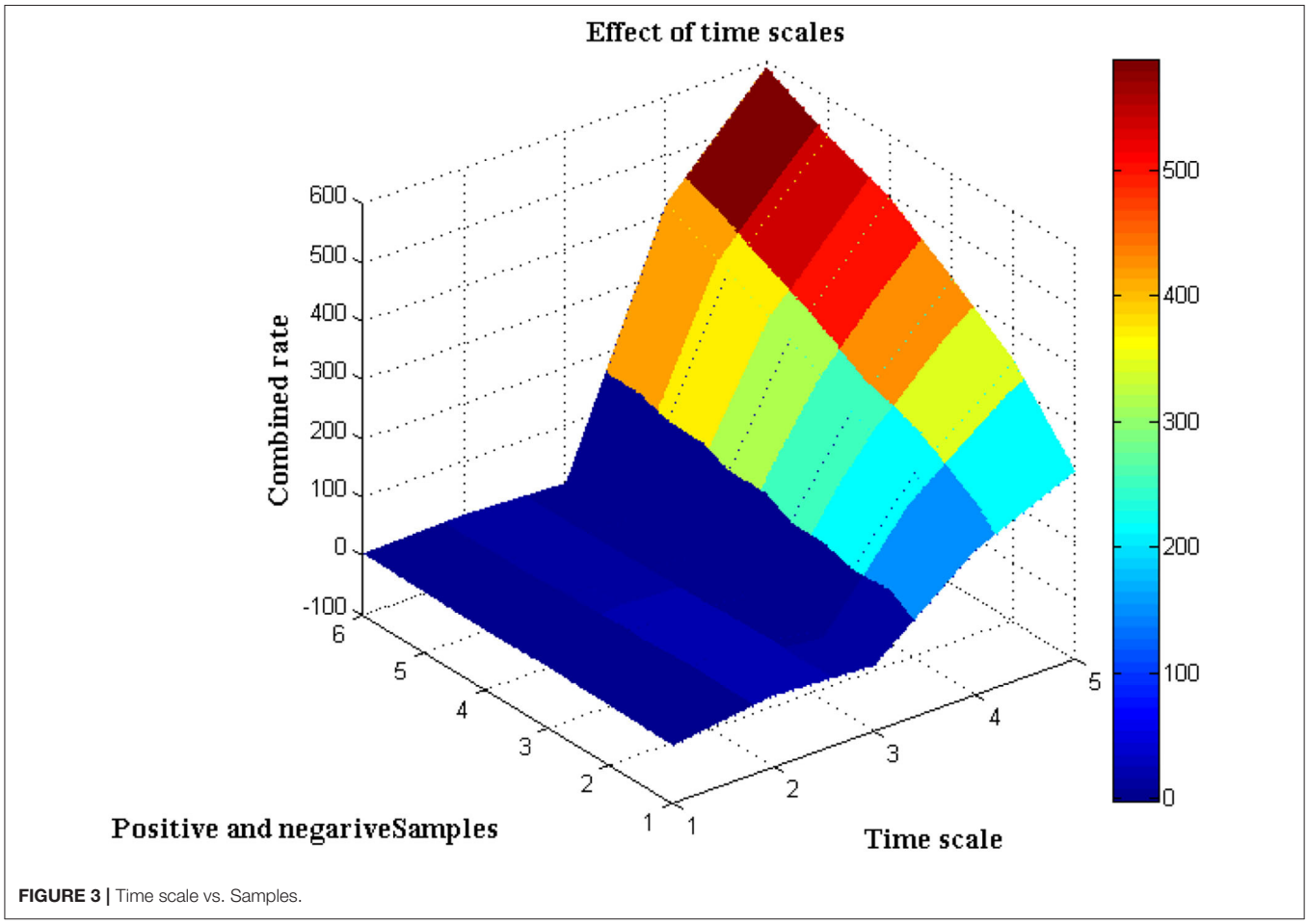


FIGURE 3 | Time scale vs. Samples.

TABLE 2 | Combined rates of samples.

Time scale	Positive samples	Negative samples	Combined rate (1)	Combined rate (Proposed)
−2	21	10	147	219
−1	24	6	213	345
0	17	3	254	423
1	13	2	314	497
2	11	2	368	536
3	8	2	416	589

and negative samples are treated as identical situations when compared to reference inputs. After three state observations, the final combined rate for both the traditional and projected methods is calculated from signal images. Once the signals have been agreed upon, the arrival rate should be matched to the time scales and maximized.

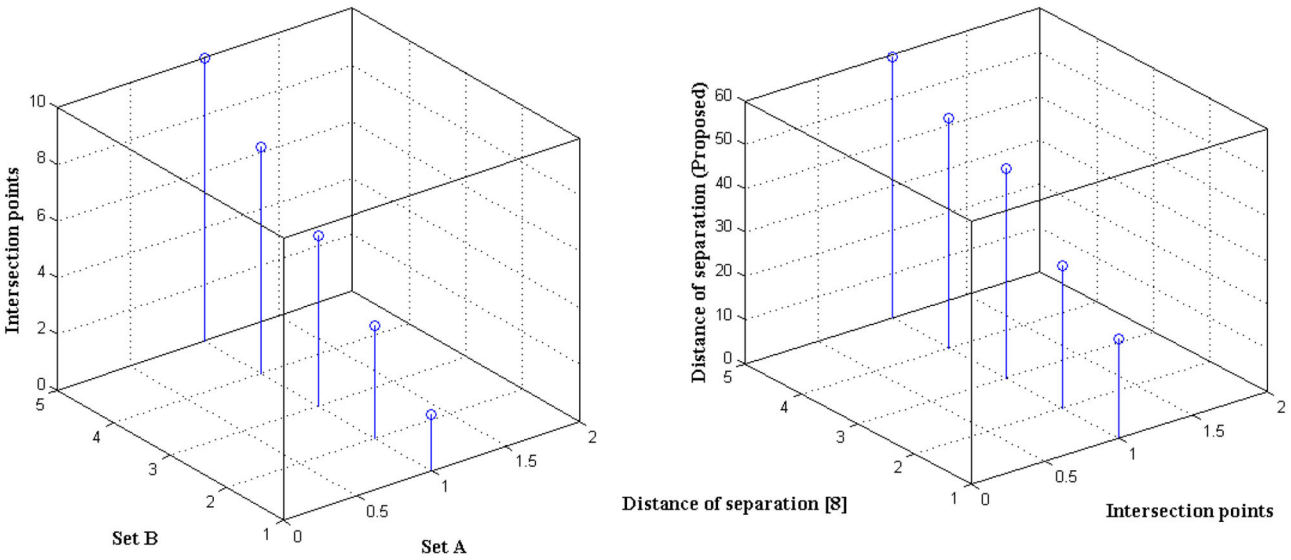


FIGURE 4 | Separation of distance in the incidence of data set.

TABLE 3 | Distance of separation.

Set 1	Set 2	Distance of separation (8)	Distance of separation (Proposed)
2	4	22.6	23.4
4	8	32.4	38.7
6	12	47.8	53.1
8	16	52.3	62.4
10	20	59.5	66.9

If the maximum rate is not obtained, it means that some stages of signal separation are not being processed, implying that loop depiction must begin at the beginning. According to the contrasted combined rate values, the projected technique employing deep learning models optimizes the rate at 589 signals per second, whereas the existing method can only maximize up to 416 signals per second, which is significantly less than the planned adaption.

Scenario 3

The distance of monitoring between numerous variations is an important case study after the representation of time scale in biological signals. The suggested technique considers four discrete samples, with the point of inter-section providing the actual distance. When the separation distance increases by a positive factor, it is referred to as scaling with regard to the time factor since the value of the signal representation changes. Furthermore, if deep learning is not used, the procedure might be defined as manual because positive variables must be multiplied by time scales. To avoid such manual procedures, a point of inter-section is chosen that has no effect on frequency variations. However, this method is limited since every 30 s inter-section interval, signal rotation is required. During this 30 s time period, over 1,000 signals are sampled within the channel without any data set augmentation. **Figure 4** depicts the imprecise distance separation values that were simulated.

Figure 4, **Table 3** show that two sets of values were chosen as input, the first of which differs by two value points and the second of which differ by four. These two sets are combined, and a point of inter-section is provided, as well as values at each midway, resulting in automatic distance configuration. For biological signals, a sensing device must be placed at each midway to maximize the separation distance. If the separation distance is not maximized, data will be augmented, which is a severe flaw in traditional models without deep learning techniques. However, the proposed system model solves it using five different points of touch, reducing the number of sensing devices from ten to five. Furthermore, 5 sensing devices are sufficient for viewing the representation signal at all phases, as evidenced by the last data set, where the suggested method achieves a distance of separation of 66.9, whilst the conventional methodology achieves a distance of separation of 59.5.

Scenario 4

This scenario is implemented utilizing the activation function, which is represented using Equation, for further analysis of the data set (13). The number of iterations considered in this sort of initiation varies for each defined data set. It is characterized

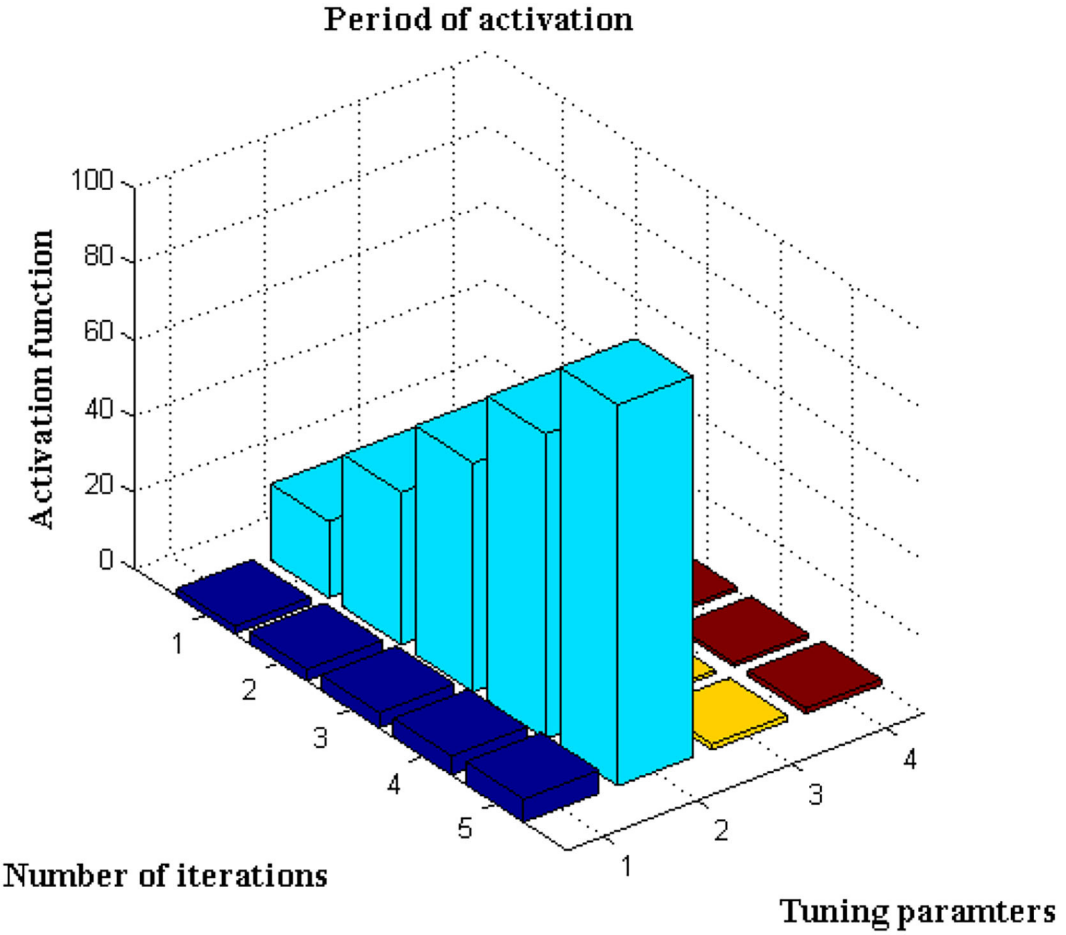


FIGURE 5 | Activation function with iteration periods.

TABLE 4 | Iteration and activation function.

Tuning parameters	Number of iterations	Activation function (9)	Activation function (Proposed)
2	20	0.5	1.5
3	40	0.5	2
4	60	1	2
5	80	1	2.5
6	100	1.5	3

as an exact network if the activation period is obtained at negligible iteration, hence the maximum iteration period is set to 100 with a variation of 20 predetermined periods. The main rationale for choosing 20 defined periods is that during this sample, error values are decreased to 0.001, resulting in the best rate of concealed and output markers. Only if the following constraints are present, such as <4k Hz for variable activities and >30kHz for constant activities, will the training function be active. If any value exceeds a specific range, it means that the visual representation's border condition is incorrect, and testing data cannot be activated. After extensive testing, it was discovered that the proposed method's boundary conditions are met, and a visual representation is obtained, as shown in **Figure 5**.

Figure 5, **Table 4** show that the bar that rises to a large extent is labeled as the activation period, whereas there is no rise in activation functions at three different stages. Aside from this activation function, the number of tuning parameters must be chosen in a ratio that is appropriate for the number of iterations. The proposed method increases the ratio to around 1:15, resulting in the best activation function within the time constraints. This is demonstrated at iteration 80, where a high activation function is achieved for 2.5 scaling periods.

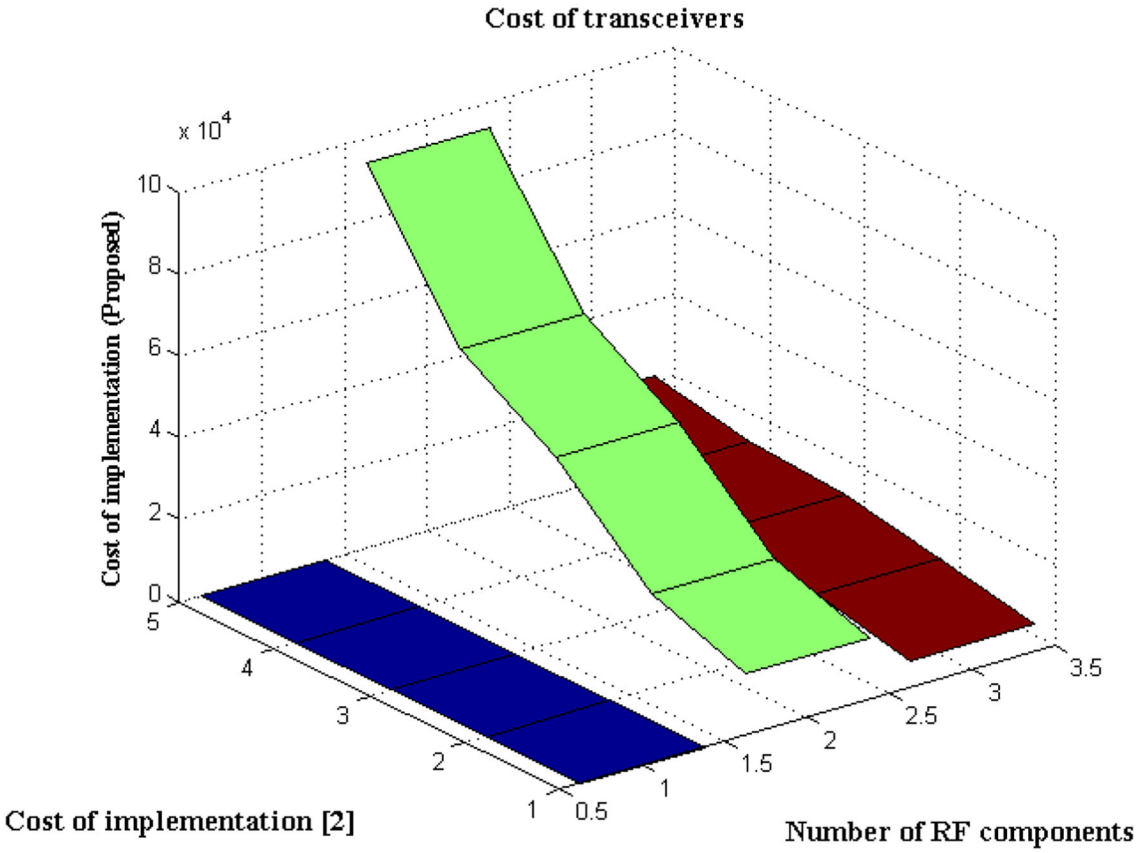


FIGURE 6 | Implementation cost.

Because the proportion ratio in the existing model (9) is not properly defined, the activation period remains at 1 for even two iteration periods. This demonstrates that the projected deep learning model produces the best iteration results with the best activation functions.

Scenario 5

This Scenario depicts the cost of implementing observation units with hidden fragments. It is self-evident that the cost of implementation is decreased due to the minimal number of sensors deployed at junctions.

The type of transceiver used in the proposed technique is determined by the communication between different receivers and the control power of the stations. A radio frequency component is used as an external controller in the transceiver, and a series of recordings are made. This radio frequency component will raise the cost of signal acquisition, but a Nanoscale module is used in the deep learning process. **Figure 6** depicts the simulated implementation cost with comparison. **Figure 6** and **Table 5** show that the cost of implementation increases as the number of radiofrequency components increases for both proposed and existing methods (2). However, with the same number of components, the existing method has a high implementation cost of INR 95000. However, the proposed method is implemented with INR21500 as the transceiver is designed properly at appropriate stages using the same components. Even the power constraint is met at the same cost of installation, resulting in cost savings. This demonstrates that the cost of real-time implementation may be significantly reduced, and medical enterprises can readily obtain it.

TABLE 5 | Cost of existing and proposed methods.

Number of RF components	Cost (2)	Cost (Proposed)
3	15000	6700
6	23000	11300
9	45000	15600
12	60000	17200
15	94000	21500

Performance Analysis

Since DCS are implemented in the proposed model it is essential to check the performance characteristics in addition to the above-mentioned scenarios. The major intention for checking the enactment is that DCS will have more than one

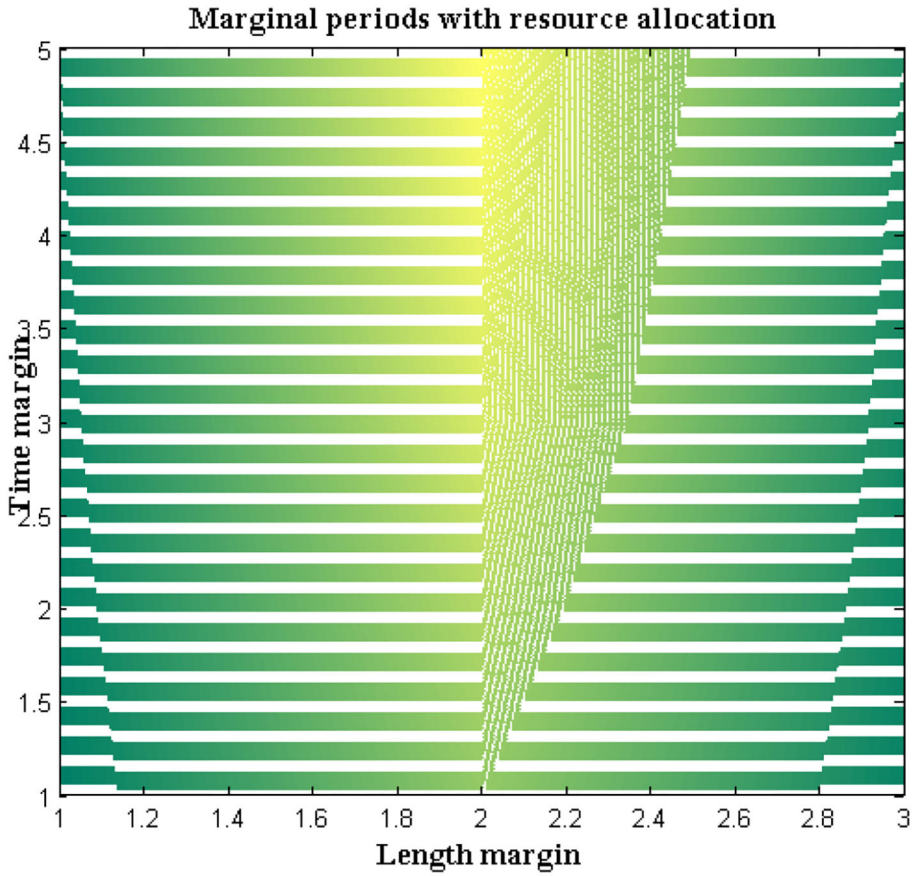


FIGURE 7 | Marginal periods with the allocation of resources.

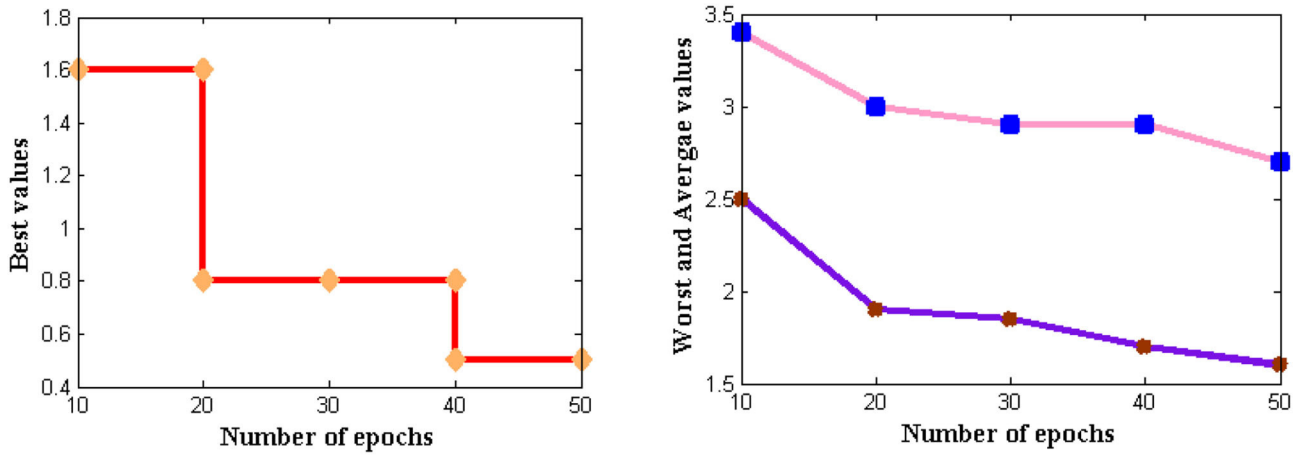


FIGURE 8 | The complexity of time intervals.

output representation and in this case, the best output must be selected and it should be compared with reference data. This output representation wills ensemble the fitness function and it should be terminated after a certain number of algorithmic steps. Also, the implemented algorithm must provide high feasibility to all system representation techniques. Therefore three different performance analyses are carried out in this section as listed below,

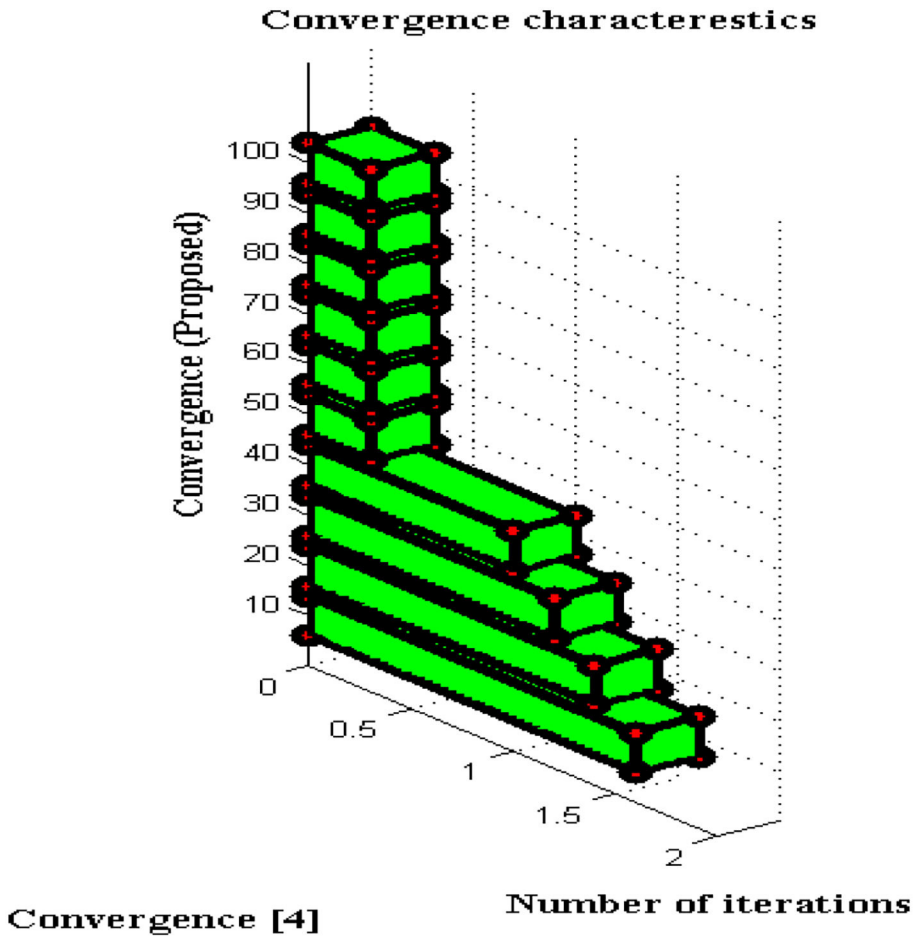


FIGURE 9 | Convergence characteristics.

Feasibility of DCS

In the proposed method a separate programming code is inscribed where step-by-step directions are included in the loop. The pre-defined loop features must be available with highly effective resources and a matching technique will be represented to check the allocation of resources at exact location points. Also, the probability distribution of a value must be defined with finite cases where both length and time margins are provided with an average value set.

For all the data sets a non-trivial solution will be achieved with two different exponential functions. Additional sub-exponential functions are also derived to determine the performance of DCS using entropy function determinations where a feasible solution can be achieved. The accomplished solutions are deliberated in **Figure 7**. From **Figure 7** it can be observed that a number of resources are varied from 4 to 20 and for each resource, both time and length margins are measured. For the proposed method both margins are kept as a minimum at the constraint for margins are satisfied as compared to reference values. However, comparisons are not made with any existing method due to unavailable resources in the selection process. Even the time margins which are represented in seconds are much lesser and the same can be applied to conjunction characteristics at border length margins. This can be proved with a maximum resource where length margins are 204 meters and, in this case, the maximum time period is 34 s. The above-mentioned values are much lesser as in the absence of resources more than 80 s time periods are observed.

Computational Complexity

If the feasibility of the proposed method is higher, then complexities that are present in time periods can be analyzed. Thus, complexities in time periods are measured in addition to memory requirements where for large input size three different values such as best, worst, and average are represented. In the projected model, the time for executing each statement is found and it is determined using the asymptotic function. The complexity in time periods is measured using several external factors such as the size of the input, processor speed, etc. The computational complexity for DCS is simulated in **Figure 8**.

From **Figure 8** it can be observed that a number of iterations is considered from 10 to 50 and all three values are plotted where

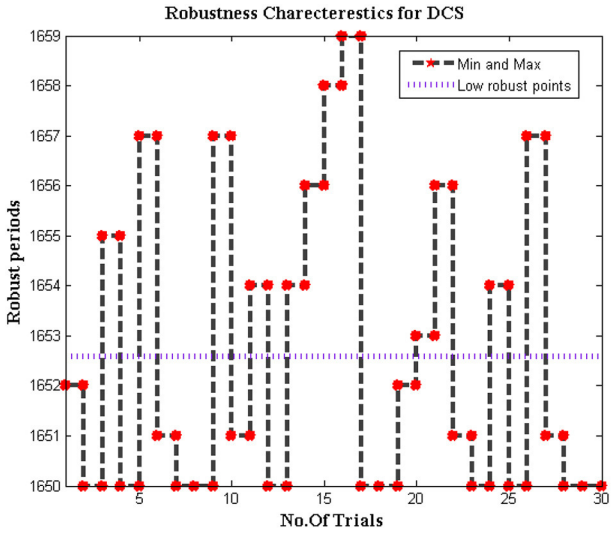


FIGURE 10 | Low robust points of DCS.

best values are shown in subplot 1, worst and average values in subplot 2. From these two subplots, DCS values are varied where low complexities are found as best values reach 0.5 s and further reduce after several iteration periods. If zero measurements are found then it indicates that time complexity is not present in the system. The same situation is observed for DCS after 80 iterations but until it reaches the corresponding iteration value complexities remain constant at 0.5 s. This proves that at constant speed DCS can able to reduce complexities in time.

Convergence Characteristics

For all non-linear system models, it is essential to perceive the convergence characteristics where persistent phases must be observed. In DCS two types of characteristics are analyzed with respect to definite and captivating problems. This will give a clear solution to address the problems that arise during a long period of time where if convergence is not achieved then that particular time period will be separated. Furthermore, the separated time periods will be analyzed in several ways such as fault communication module, signal transmission properties, sensing capabilities, etc. If any one parameter fails to provide precise results then convergence cannot be achieved and it indicates failure of the model. The convergence characteristics of DCS are deliberated in **Figure 9**.

From **Figure 9** it can be perceived that DCS is converged immediately at 50 iteration periods whereas the existing system converges only at 80 iteration periods. Therefore a large difference in convergence can be observed in the value determinations such as 0.3 and 0.5. Even after attaining fast convergence, the proposed method can be able to reduce the period of signal transfer to the receiver. This in turn helps the users to get a high signaling rate at shorter periods but fabrication parameters must have concurred with DCS. With the fabrication parameters, it is much easier to install a signal processing sensor thereby making DCS converged at initial point conditions.

Robustness of DCS

Since the dataset of the proposed method is varied as different bio-medical signals are transferred to end systems. In addition, the biomedical signals that are passed from individuals are much different when it is compared with distinct parameters. Therefore for change in signal representation, it is essential to check whether the system is robust to all input conditions. Further error in DCS must be reduced to increase the effectiveness of the proposed model as the trained inputs must achieve much closer outcomes during deduction segments. Thus the simulated model for checking the robustness is plotted in **Figure 10**.

From **Figure 10** it can be observed that the number of iterations varied from 10 to 100 and for each representation, a midpoint robust values are achieved. These midpoint values are considered as the initial design of the biomedical signal representation must not be changed as local points will provide information about missing values in the system. For all changes in iteration periods, the proposed method operates under low robust conditions whereas the conventional systems are highly robust against input conditions as local points cannot able to identify distinct data set points.

CONCLUSIONS

This paper examines a real-time augmentation biomedical signal that is used in the testing and classification mechanism of a human interface system. The application of biomedical signals in medical applications provides clear insights into future developments under the automatic mode of operation in a wide range of research fields. As a result, the proposed method prepares an automatic process using a sensing device that can be accessed in simple environmental conditions. However, a few sensors that monitor biomedical signal parameters must be fabricated in order to provide precise data and transmit signal data to the control center. In the proposed method, a pre-processing technique is used at stage 1 to classify and train the images, with the result that any incongruent signals will be detected quickly. In addition, for the training set, the previous data set has been integrated, and a comparative analysis has been performed using two algorithms: firefly and gate count technology. Since biomedical signals in the transceiver are detected using a deep learning algorithm, which includes DCS with several learning parameters in the 2.5-time scale period range. As a result, the generated data set is observed for three discrete periods using weight transformations, resulting in deep image representation extensions.

Furthermore, real-time simulation results show that biomedical signals are activated without any peripheral intrusion, even after several different classifications. To gain a better understanding of DCS, five different scenarios are divided and tested using a hardware setup that yields effective results. Biomedical signals can be used as a

building block for all knowledge transformation techniques in the future, resulting in a significant reduction in data augmentation measures.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary files, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Data curation: AY. Writing-original draft: HM. Supervision: SS. Project administration, conceptualization, and visualization: SS and HM. Methodology and resources: SS and MU. Validation: HM, HA, and AY. Review, editing, and funding acquisition: C-LC and C-MW. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China (No. 51808474) and the Ministry of Science and Technology in Taiwan (Nos. MOST 110-2218-E-305-001–MBK and MOST 110-2410-H-324-004-MY2).

ACKNOWLEDGMENTS

We wish to thank our funder C-LC who is supported by the Ministry of Science and Technology, Taiwan.

REFERENCES

1. Dahrouj H, Alghamdi R, Alwazani H, Bahanshal S, Ahmad AA, Faisal A, et al. An overview of machine learning-based techniques for solving optimization problems in communications and signal processing. *IEEE Access.* (2021) 9:74908–38. doi: 10.1109/ACCESS.2021.3079639
2. Rajeswari J, Jagannath M. Advances in biomedical signal and image processing – a systematic review. *Informatics Med Unlocked.* (2017) 8:13–9. doi: 10.1016/j.imu.2017.04.002
3. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep learning and its applications in biomedicine. *Genomics, Proteomics Bioinforma.* (2018) 16:17–32. doi: 10.1016/j.gpb.2017.07.003
4. Zemouri R, Zerhouni N, Racoceanu D. Deep learning in the biomedical applications: Recent and future status. *Appl Sci.* (2019) 9. doi: 10.3390/app9081526
5. Belo D, Rodrigues J, Vaz JR, Pezarat-Correia P, Gamboa H. Biosignals learning and synthesis using deep neural networks. *Biomed Eng Online.* (2017) 16:1–17. doi: 10.1186/s12938-017-0405-0
6. Nayak J, Naik B, Dinesh P, Vakula K, Dash PB. Firefly algorithm in biomedical and health care: advances, issues and challenges. *SN Comput Sci.* (2020) 1:311. doi: 10.1007/s42979-020-00320-x
7. Ahmed ST, Sandhya M, Sankar S. An optimized RTSRV machine learning algorithm for biomedical signal transmission and regeneration for telemedicine environment. *Procedia Comput Sci.* (2019) 152:140–9. doi: 10.1016/j.procs.2019.05.036
8. Hu T, Khishe M, Mohammadi M, Parvizi GR, Taher Karim SH, Rashid TA. Real-Time COVID-19 diagnosis from X-Ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm. *Biomed Signal Process Control.* (2021) 68. doi: 10.1016/j.bspc.2021.102764
9. Jain N, Mishra B, Wilson P. A Low gate count reconfigurable architecture for biomedical signal processing applications. *SN Appl Sci.* (2021) 3:1–19. doi: 10.1007/s42452-021-04412-y
10. Alaskar H. Convolutional neural network application in biomedical signals. *J Comput Sci Inf Technol.* (2018) 6:45–59. doi: 10.15640/jcsit.v6n2a5
11. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: a systematic review. *J Neural Eng.* (2019) 16. doi: 10.1088/1741-2552/ab260c
12. Konstantinidis EI, Frantzidis CA, Tzimkas L, Pappas C, Bamidis PD. Accelerating biomedical signal processing algorithms with parallel programming on graphic processor units. In: *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine*. Larnaca: IEEE (2009). doi: 10.1109/ITAB.2009.5394314
13. De Jonckheere J, Logier R, Dassonneville A, Delmar G, Vasseur C. PhysioTrace: an efficient toolkit for biomedical signal processing. *Annu Int Conf IEEE Eng Med Biol - Proc 7 VOLS.* (2005) 6739–41. doi: 10.1109/IEMBS.2005.1616051
14. Stojanović R, KneŽević S, Karadaglić D, DevedŽić G. Optimization and implementation of the wavelet based algorithms for embedded biomedical signal processing. *Comput Sci Inf Syst.* (2013) 10:503–23. doi: 10.2298/CSIS120517013S
15. Pandya S, Gadekallu TR, Reddy PK, Wang W, Alazab M. Infused heart: a novel knowledge-infused learning framework for diagnosis of cardiovascular events. *IEEE Trans Computational Soc Sys.* (2022) 17:1–12. doi: 10.1109/TCSS.2022.3151643
16. Arikumar KS, Prathiba SB, Alazab M, Gadekallu TR, Pandya S, Khan JM, et al. FL-PMI: federated learning-based person movement identification through wearable devices in smart healthcare ystems. *Sensors.* (2022) 22. doi: 10.3390/s22041377
17. Saleem K, Saleem M, Zeeshan R, Javed AR, Alazab M, Gadekallu TR, et al. Situation-aware BDI reasoning to detect early symptoms of Covid 19 using smartwatch. *IEEE Sensors J.* (2022) 1–8. doi: 10.1109/JSEN.2022.3156819
18. Fokas AS, Dikaios N, Kastis GA. Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. *J R Soc Interface.* (2020) 17. doi: 10.1098/rsif.2020.0494
19. Iglesias-Puzas, Boixeda, P. Deep learning and mathematical models in dermatology. *Actas Dermosifiliogr.* (2020) 111:192–95. doi: 10.1016/j.ad.2019.01.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Manoharan, Selvarajan, Yafoz, Alterazi, Uddin, Chen and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fpubh.2022.885624

A Visualized Dynamic Prediction Model for Overall Survival in Elderly Patients With Pancreatic Cancer for Smart Medical Services

Jiang Zhong[1†], XingShu Liao[1†], Shuang Peng[2], Junyi Cao[3], Yue Liu[4], Chunyang Liu[5], Ju Qiu[5], Xiaoyan Guan[4], Yang Zhang[6], Xiaozhu Liu[7*] and Shengxian Peng[5*]*

[1] College of Computer Science, Chongqing University, Chongqing, China, [2] General Affairs Section, The People's Hospital of Tongnan District, Chongqing, China, [3] Department of Medical Quality Control, First People's Hospital of Zigong City, Zigong, China, [4] Department of Pediatrics, First People's Hospital of Zigong City, Zigong, China, [5] Scientific Research Department, First People's Hospital of Zigong City, Zigong, China, [6] College of Medical Information, Chongqing Medical University, Chongqing, China, [7] Department of Cardiology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Shahab S. Band,
National Yunlin University of Science and Technology, Taiwan
Praise Young,
Nnamdi Azikiwe University, Nigeria
Shweta Agrawal,
Sage University, India

***Correspondence:**
Yang Zhang
1565983618@qq.com
Xiaozhu Liu
xiaozhuliu2021@163.com
Shengxian Peng
13258280319@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 28 February 2022
Accepted: 06 April 2022
Published: 24 May 2022

Citation:
Zhong J, Liao X, Peng S, Cao J, Liu Y, Liu C, Qiu J, Guan X, Zhang Y, Liu X and Peng S (2022) A Visualized Dynamic Prediction Model for Overall Survival in Elderly Patients With Pancreatic Cancer for Smart Medical Services. Front. Public Health 10:885624. doi: 10.3389/fpubh.2022.885624

Background: Pancreatic cancer (PC) is a highly malignant tumor of the digestive system. The number of elderly patients with PC is increasing, and older age is related to a worse prognosis. Accurate prognostication is crucial in treatment decisions made for people diagnosed with PC. However, an accurate predictive model for the prognosis of these patients is still lacking. We aimed to construct nomograms for predicting the overall survival (OS) of elderly patients with PC.

Methods: Patients with PC, older than 65 years old from 2010 to 2015 in the Surveillance, Epidemiology, and End Results database, were selected and randomly divided into training cohort (n = 4,586) and validation cohort (n = 1,966). Data of patients in 2016–2018 (n = 1,761) were used for external validation. Univariable and forward stepwise multivariable Cox analysis was used to determine the independent prognostic factors. We used significant variables in the training set to construct nomograms predicting prognosis. The performance of the models was evaluated for their discrimination and calibration power based on the concordance index (C-index), calibration curve, and the decision curve analysis (DCA).

Results: Age, insurance, grade, surgery, radiation, chemotherapy, T, N, and American Joint Commission on Cancer were independent predictors for OS and thus were included in our nomogram. In the training cohort and validation cohort, the C-indices of our nomogram were 0.725 (95%CI: 0.715–0.735) and 0.711 (95%CI: 0.695–0.727), respectively. The 1-, 3-, and 5-year areas under receiver operating characteristic curves showed similar results. The calibration curves showed a high consensus between observations and predictions. In the external validation cohort, C-index (0.797, 95%CI: 0.778–0.816) and calibration curves also revealed high consistency between observations and predictions. The nomogram-related DCA curves showed better clinical utility compared to tumor-node-metastasis staging. In addition, we have developed an online prediction tool for OS.

Conclusions: A web-based prediction model for OS in elderly patients with PC was constructed and validated, which may be useful for prognostic assessment, treatment strategy selection, and follow-up management of these patients.

Keywords: nomogram, elderly patients, pancreatic cancer, SEER database, online application

INTRODUCTION

Pancreatic cancer (PC) is one of the most dangerous malignancies in the world and is a lethal disease with dismal survival rates, with a 5-year overall survival rate of <5% and a median survival of ∼6–8 months after diagnosis. It is the fourth most common cause of cancer death in both men and women due to the difficulty of early detection and poor prognosis (1–3). In addition, PC is the main disease of the elderly population, as aging intensifies, the number of elderly with PC is expected to rise in the future. At the same time, it is difficult for physicians to stage and predict the prognosis of PC because it exhibits a variety of tumor behaviors and symptoms (4). Despite the continuous introduction of new surgical techniques and medical therapies, and the evolving concept of integrated management of PC, the prognosis of PC patients has improved modestly, but the 5-year survival rate has not improved significantly (5). So the poor prognosis of PC remains a major challenge for mankind.

The tumor-node-metastasis (TNM) staging system of the American Joint Commission on Cancer (AJCC) 8th edition, is commonly used for prognostic evaluation of pancreatic ductal adenocarcinoma. However, researchers have found the staging system inadequate (6–8). It only incorporates some features of the tumors into the staging system, and its prognostic factors go far beyond these. Shi et al. constructed a nomogram for predicting the overall survival (OS) and cancer-specific survival (CSS) of young patients with PC and found that primary site, pathological types, AJCC stage, and surgery were independent factors affecting CSS (8). Although many studies have confirmed that the clinical characteristics and prognosis of older patients with PC are different from young patients (9–11), the nomogram of PC in previous studies cannot accurately predict the OS of elderly patients with PC (12–14). Therefore, technically feasible and easily clinically accessible nomograms that predict the OS specifically for elderly patients with PC are still urgently required.

At present, nomograms have been developed and proposed as an innovative alternative tool for prognostic assessment of many cancers (15–17), which can combine important demographic and clinicopathological features to estimate individual survival for patients with cancer. A nomogram for patients with PC over 65 years old or older derived from population-based data, to the best of our knowledge, has never been reported. The Surveillance, Epidemiology, and End Results (SEER) database serves as the definitive cancer database in the United States. It contains data related to cancer prognosis such as age, race, marital status, histologic grade, tumor size, surgery, radiation, and chemotherapy (18). Based on the SEER database, we aim to investigate the risk factors for OS in elderly patients with PC and establish a web-based prediction model for predicting OS, which might be helpful for the prognostic prediction, and treatment strategy selection. It also can help doctors and patients to make follow-up decisions.

PATIENTS AND METHODS

Patients and Variables Inclusion

The clinicopathological data of all patients with PC from 2010 to 2018 were downloaded from the SEER database. We put the original data in **Supplementary Material** and put the relevant code on GitHub, see the link for details https://github.com/xiaoyang11223/pancreatic-cancer.git. It collects information on patient demographics, year of diagnosis, marital status, tumor size, histopathological grade, TNM stage, surgery, radiotherapy, chemotherapy, post-operative AJCC 7th staging, and follow-up for survival. The inclusion criteria were as follows: (1) age≥65years; (2) International Classification of Diseases for Oncology, 3rd edition [ICD-O-3] code 8150, 8151, 8152, 8154, 8155; (3) those with a confirmed pathological diagnosis from 2010 to 2018; and (4) those with unknown data about the grade. The exclusion criteria were as follows: (1) unknown diagnostic confirmation; (2) unknown marital status; (3) unknown insurance recode; (4) survival time <1 month; (5) tumor size≥990; and (6) unknown whether radiotherapy was performed. Finally, 6,552 eligible elderly patients with PC were included in this study. The following variables were analyzed: age, insurance, grade, surgery, radiation, chemotherapy, TNM stage, and AJCC. The detailed flow chart was shown in **Figure 1**. For this study, we have signed authorization and received permission from SEER to access and use the SEER information. The SEER database is publicly available and the data for all patients are deidentified, so institutional review board approval and informed consent were not required in this study.

Statistical Analysis

To construct and validate the nomogram and ensure the robustness of the prediction model, we divided the data into the training set and verification set according to 7:3 through the SAMPLE function of R. we randomly assigned 70% ($n = 4,586$) of patients from 2010 to 2015 and 30% ($n = 1,966$) of patients from 2010 to 2015 to the training and validation cohorts, respectively. A total of 1,761 patients in the SEER database from 2016 to 2018 were included in the external validation cohort. A chi-square test was used to compare the clinicopathological characteristics between the training set and the validation set. OS referred to the duration of PC from diagnosis to death, and patients who were alive at the point of the last follow-up were considered

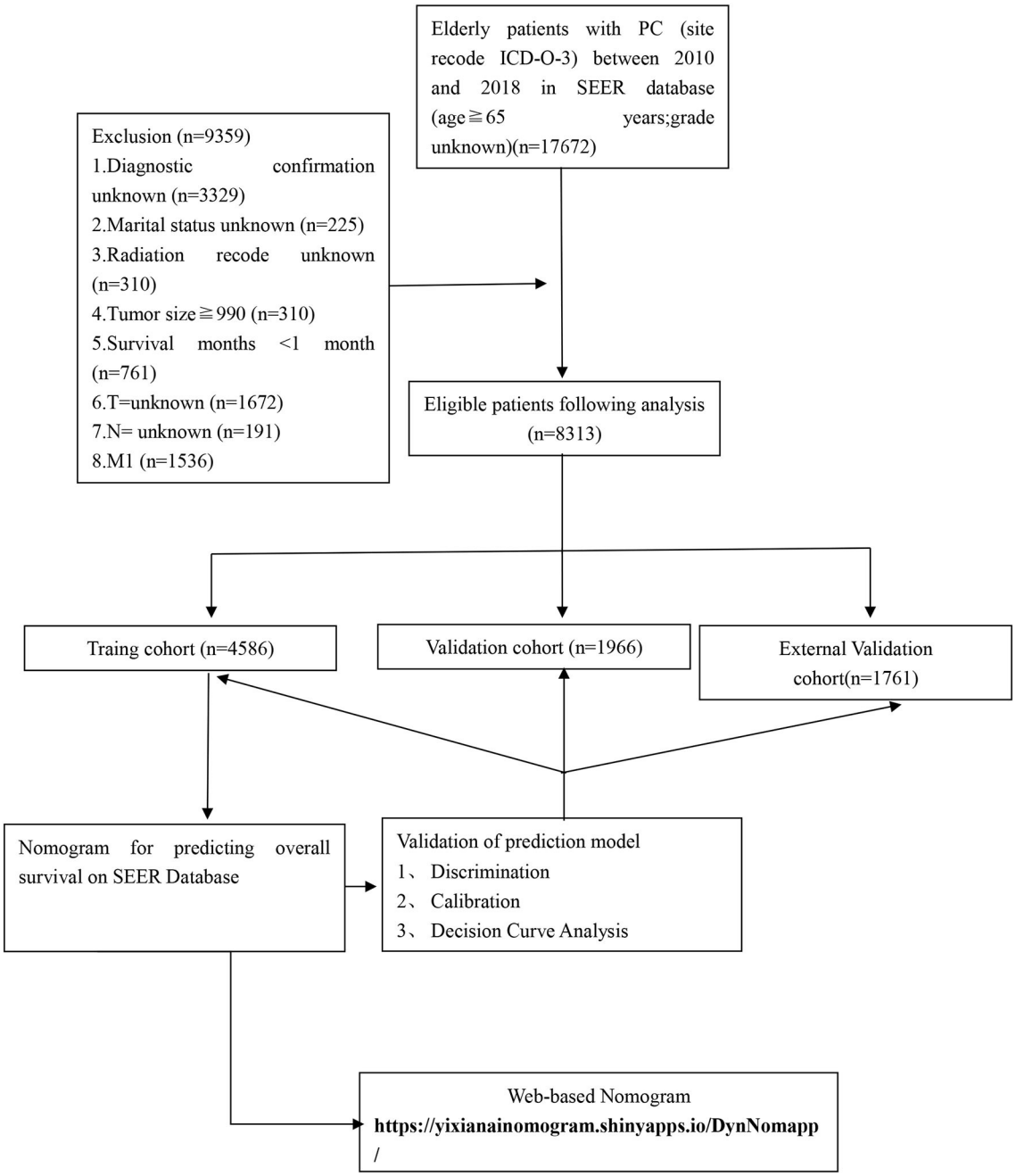


FIGURE 1 | The flowchart of including and dividing patients. Github: https://github.com/xiaoyang11223/pancreatic-cancer.git.

censored events. Variables related to OS for which $p<0.05$ in the univariable analysis or with important clinical value were entered into a multivariate logistic regression model based on proportional subdistribution hazard models to determine risk factors independently associated with the development of PC in elderly patients, which was performed using a Cox proportional hazards regression model. Discrimination of the nomograms was measured by the concordance index (C-index) and receiver operating characteristic (ROC) curve with its respective 95% CI, which quantifies the level of consistency between the observed OS and the predicted OS probability.

Clinical prediction models can provide doctors and patients with quantified risk value based on current health status to predict future health status utilizing non-invasive, low-cost, and easily collected indicators, which has important health economics significance. Decision curve analysis (DCA) is a new calculation method that estimates the net benefits under various risk thresholds to evaluate the clinical value of the

model. We used DCA to evaluate the clinical practicability of our nomogram. Based on the cut-off value calculated from the total score of our nomogram, we divided patients into a low-risk group and a high-risk group. We applied Kaplan–Meier curves and log-rank tests to compare patient survival between different groups. The nomogram, C-indices, ROCs, calibration curves, DCA curves, Kaplan–Meier curves, and a web application for survival prediction were by utilizing R software. For all statistical analyses, p-value < 0.05 was considered statistically significant.

RESULTS

Patients Characteristic

Our study flowchart is shown in **Figure 1**. Finally, there were a total of 6,552 cases of elderly patients with PC diagnosed in the SEER database between 1 January 2010 and 31 December 2015. These patients were randomly divided into a training set (N =4,586) and a validation set (N = 1,966). The demographic and clinicopathological information of patients were listed in **Table 1**. Of these patients, 3,702 cases (56.50%) were 65–74 years old, 2,850 cases (43.50%) were >74 years old, 5,501 (83.96%) were white, 3,284 (50.12%) were male, 5,945 (90.74%) were married, and 4,683 (71.47%) were T3 stage. The histological tumor grades were 1,395 (21.29%), 22,839 (43.33%), 2,209 (33.71%), and 109 (1.66%) for patients with tumor histological grades I, II, III, and IV, respectively. Their AJCC stages were IA: n=640 (9.77%), IB: n=823 (12.56%), IIA: n=1,724 (26.31%), and IIB: n=3365 (51.36%), respectively. There were 3,348 (51.10%) underwent local or pancreatectomy, 3,649 (55.69%) underwent chemotherapy, and 5,043 (76.97%) did not receive radiotherapy. Detailed patient clinical characteristics were summarized in **Table 1**. There were no significant differences in clinicopathologic characteristics between the training and the validation set. Characteristics of patients in the external validation cohort are shown in **Supplementary Table S1**.

Univariable and Multivariable Cox Regression Analysis

Univariate Cox regression analysis was used to screen the prognostic factors of PCC in elderly patients, namely, sex, age, insurance, Fuhrman grade, surgery, radiation, chemotherapy, T stage, N stage, and AJCC stage (**Table 2**). Then we developed multivariate Cox models using the selected factors to identify independent risk factors. The hazard ratio (HR) is used to quantify its effect on OS. Multivariate analysis found that age, insurance, grade, T stage, N stage, surgery, radiation, chemotherapy, and AJCC stage were significant independent risk factors for the prognosis of patients. The results of univariate and multivariate factors are shown in **Table 2**.

Nomogram Construction for 1-, 3-, and 5-Year OS

The multivariate Cox regression analysis identified the independent risk factors used to construct a nomogram to predict elderly patients with PC at 1-, 3-, and 5-year OS (**Figure 2**). The nomogram indicated that surgery and grade are still the most significant factors affecting the prognosis of patients, followed by age, T stage, AJCC stage, and chemotherapy.

Validation of the Nomogram

The C-index of the training set and the validation set were 0.725 (95% CI: 0.715–0.735) and 0.711 (95% CI: 0.695–0.727), respectively, indicating that the nomogram has good discriminatory power. In the external validation cohort, C-index was 0.797 (95% CI: 0.778–0.816). The calibration plots for the training cohort and the validation cohort used to predict OS show an excellent consistency between the actually observed and nomogram-predicted survival (**Figures 3A–C**). In the training cohort, the AUC of the predicted nomogram for 1-, 3-, and 5-year were 0.787, 0.766, and 0.35 (**Figure 4A**). In the validation cohort, the AUC of the predicted nomogram for 1-, 3-, and 5-year were 0.771, 0.751, and 0.777 (**Figure 4B**), respectively. Furthermore, most patients do not survive longer than 35 months. In the external validation, the AUC of the predicted nomogram for 1 year was 0.828 (**Figure 4C**).

Clinical Application of the Nomogram

Decision curve analyses (DCAs) showed that the clinical application value of nomograms was superior to that of the traditional TNM stage system (**Figures 5A,B**). In addition, we constructed a risk stratification system based on the total points of patients on a nomogram. All patients were divided into a risk-low group (overall score ≤185.1) and a risk-high group (total score >185.1). The Kaplan–Meier curves of the validation and training sets demonstrated that there were significant differences in the survival of patients in each risk group (**Figure 6**). That is to say, Kaplan–Meier curve results well-proved the discriminant ability of the nomogram prediction model.

Online Application for OS Prediction

Based on this nomogram, we have developed a network calculator to predict the OS of elderly patients with PC, which can be accessed at https://yixianainomogram.shinyapps.io/DynNomapp/. Enter the clinicopathological characteristics of the patient, you can immediately get the predictive survival probability of the patient. In conclusion, this online prediction tool is simple, easy to understand, and easy to use in clinical practice.

DISCUSSION

In this study, we used abundant clinical samples to establish a nomogram to effectively predict 1-, 3-, and 5-year OS in elderly patients with PC based on the SEER database. At the same time, we build the nomogram model and validate the performance of the model by screening clinically significant variables. We determined that surgery, grade, radiation, T stage, N stage, AJCC stage, chemotherapy, age, and insurance were independent factors affecting the OS of elderly patients with PC by univariable and multivariable regression analysis based on the SEER database. Furthermore, a combination of key elements of traditional staging systems and other tumor-associated indicators, such as age, grade, surgery, and

TABLE 1 | Demographical and clinical characteristics of training set and validation set.

Characteristics	All patients *N* = 6,552	Training cohort *N* = 4,586	Validation cohort *N* = 1,966	*P*
Age (%)				0.2609
65–74	3,702 (56.50)	2,570 (56.04)	1,132 (57.58)	
>74	2,850 (43.50)	2,016 (43.96)	834 (42.42)	
Sex (%)				0.48
Female	3,268 (49.88)	2,301 (50.17)	967 (49.19)	
Male	3,284 (50.12)	2,285 (49.83)	999 (50.81)	
Race (%)				<0.001
Black	503 (7.68)	283 (6.17)	220 (11.19)	
Other[a]	548 (8.36)	495 (10.79)	53 (2.70)	
White	5,501 (83.96)	3,808 (83.04)	1,693 (86.11)	
Marital state (%)				0.24
Married	5,945 (90.74)	4,148 (90.45)	1,797 (91.40)	
No	607 (9.26)	438 (9.55)	169 (8.60)	
Insurance (%)				0.0003
Insured	6,056 (92.43)	4,203 (91.65)	1,853 (94.25)	
No	496 (7.57)	383 (8.35)	113 (5.75)	
Grade (%)				0.0834
I	1,395 (21.29)	975 (21.26)	420 (21.36)	
II	2,839 (43.33)	2,016 (43.96)	823 (41.86)	
III	2,209 (33.71)	1,511 (32.95)	698 (35.50)	
IV	109 (1.66)	84 (1.83)	25 (1.27)	
T (%)				0.0025
T1	709 (10.82)	516 (11.25)	193 (9.82)	
T2	1,160 (17.70)	850 (18.53)	310 (15.77)	
T3	4,683 (71.47)	3,220 (70.21)	1,463 (74.42)	
N (%)				0.0535
N0	3,187 (48.64)	2,267 (49.43)	920 (46.80)	
N1	3,365 (51.36)	2,319 (50.57)	1,046 (53.20)	
AJCC (%)				0.0833
IA	640 (9.77)	459 (10.01)	181 (9.21)	
IB	823 (12.56)	602 (13.13)	221 (11.24)	
IIA	1,724 (26.31)	1,206 (26.30)	518 (26.35)	
IIB	3,365 (51.36)	2,319 (50.57)	1,046 (53.20)	
Surgery (%)				<0.001
Extended pancreatoduodenectomy	219 (3.34)	122 (2.66)	97 (4.93)	
Local excision of tumor	1,121 (17.11)	777 (16.94)	344 (17.50)	
Local or pancreatectomy	3,348 (51.10)	2,357 (51.40)	991 (50.41)	
No	1,264 (19.29)	936 (20.41)	328 (16.68)	
Total pancreatectomy	600 (9.16)	394 (8.59)	206 (10.48)	
Radiation (%)				0.0001
No	5,043 (76.97)	3,591 (78.30)	1,452 (73.86)	
Yes	1,509 (23.03)	995 (21.70)	514 (26.14)	
Chemotherapy (%)				0.8526
No	2,903 (44.31)	2,028 (44.22)	875 (44.51)	
Yes	3,649 (55.69)	2,558 (55.78)	1,091 (55.49)	
Tumor size, mm [median (IQR)]		31.000 [24.000, 41.000]	31.000 [24.000, 41.000]	0.7675

[a] *Other includes American Indian/AK Native, Asian/Pacific Islander.*

chemotherapy, ensures that the nomograms display excellent discriminant ability in predicting OS compared with 8th edition TNM staging systems, which can be seen from the comparison of c-index and values of AUC (19, 20).

Nomogram, a simple statistical predictive tool, has been constructed in PC previously and proved to be useful and effective. Several nomograms have been constructed in patients with PC, and show more accurate survival prediction than the

TABLE 2 | Univariable and multivariable Cox regression analysis of OS in training set.

Variable	Univariate analysis		Multivariate analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Sex				
Female	Reference		Reference	
Male	0.94 (0.88–0.99)	0.029	1.01 (0.96–1.08)	0.662
Race				
Black	Reference		Reference	
Other	0.88 (0.76–1.02)	0.095	0.9 (0.78–1.05)	0.171
White	0.91 (0.81–1.01)	0.08	0.95 (0.85–1.06)	0.319
Age				
>74	Reference		Reference	
65–74	0.70 (0.66–0.74)	<0.001	0.81 (0.76–0.86)	<0.001
Insurance				
Insured	Reference		Reference	
No	1.24 (1.12–1.39)	<0.001	1.13 (1.01–1.26)	0.030
Marital.state				
Married	Reference		Reference	
No	1.09 (0.99–1.21)	0.08	1.07 (0.97–1.19)	0.196
Grade				
I	Reference		Reference	
II	2.16 (1.97–2.37)	<0.001	2.1 (1.9–2.31)	<0.001
III	3.14 (2.86–3.45)	<0.001	2.9 (2.62–3.21)	<0.001
IV	2.85 (2.26–3.59)	<0.001	2.48 (1.96–3.14)	<0.001
Surgery				
Extended pancreatoduodenectomy	Reference		Reference	
Local excision of tumor	0.61 (0.51–0.73)	<0.001	0.94 (0.78–1.13)	0.485
Local or pancreatectomy	1.01 (0.85–1.19)	0.912	1.12 (0.95–1.33)	0.178
No	2.60 (2.19–3.09)	<0.001	3.88 (3.25–4.63)	<0.001
Total pancreatectomy	0.99 (0.82–1.19)	0.889	1.13 (0.93–1.36)	0.210
Radiation				
No	Reference		Reference	
Yes	0.91 (0.85–0.98)	0.01	0.86 (0.8–0.93)	<0.001
Chemotherapy				
No	Reference		Reference	
Yes	0.94 (0.89–1)	0.056	0.64 (0.59–0.68)	<0.001
T				
T1	Reference		Reference	
T2	2.73 (2.35–3.16)	<0.001	1.49 (1.05–2.12)	0.025
T3	3.47 (3.04–3.96)	<0.001	1.89 (1.36–2.63)	<0.001
N				
N0	Reference		Reference	
N1	1.40 (1.32–1.49)	<0.001	2.07 (1.45–2.96)	<0.001
AJCC				
IA	Reference		Reference	
IB	2.72 (2.31–3.21)	<0.001	1.33 (0.91–1.97)	0.145
IIA	3.52 (3.03–4.09)	<0.001	1.51 (1.05–2.17)	0.026
IIB	3.71 (3.21–4.29)	<0.001	0.73 (0.68–0.79)	<0.001

conventional staging system in different populations (8, 13, 14). Although many studies have confirmed that the clinical characteristics and prognosis of older patients with PC are different from young patients (9–11), the nomogram of PC in previous studies cannot accurately predict the OS of elderly patients with PC (12–14). Compared with patients with PC of

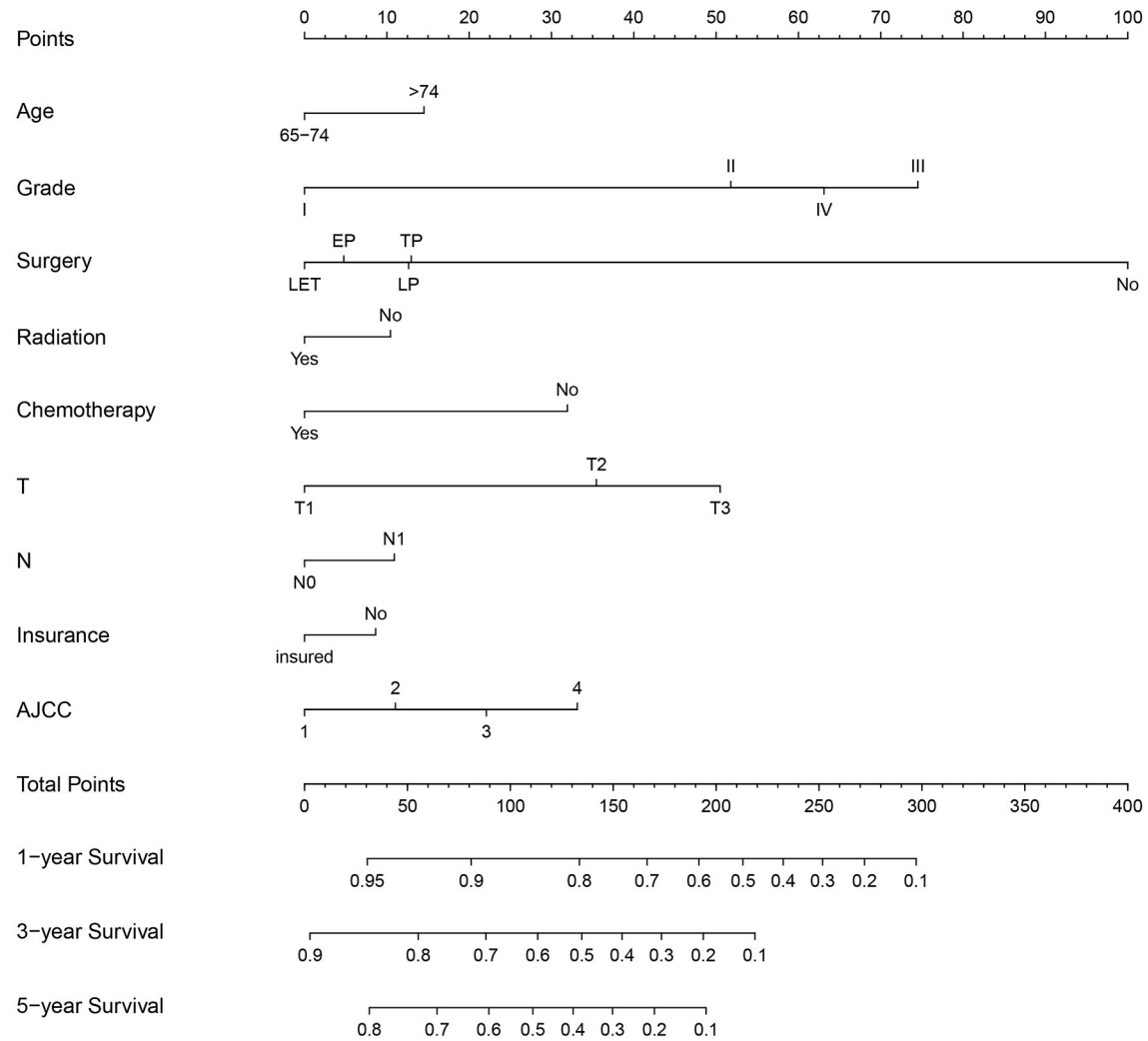


FIGURE 2 | Nomogram for 1-, 3-, and 5-year OS of elderly patients with PC.

younger ages, elderly patients with PC have unique physiological characteristics. These characteristics include more comorbidities, less life expectancy, and worse life expectancy. Therefore, it was clear that there was an urgent need to develop predictive models for elderly patients with PC. Previous studies have found that age was an independent factor affecting OS in patients with PC (12–14, 21). Shi et al. (8) have found that the nomograms could effectively predict OS and CSS in young patients with PC, which helps clinicians more accurately and quantitatively judge the prognosis of individual patients. Li and Liu (13) have found that their nomogram for OS predicting can serve as an efficacious survival-predicting model and assist in accurate decision-making for patients over 40 years old with surgically resected PC. The clinicopathological features and prognosis of young patients with PC are not quite the same as those of elderly patients with PC (9–11, 22, 23). Another study used several variables of age, differentiation, TNM stage, surgery, and lymph node surgery to construct a nomogram predicting OS rates in patients with PC having age groups of 25–39, 40–59, 60–79, and 80+ (8). But the study still had some restrictions, they did not include patients who have undergone chemotherapy or radiotherapy. To our knowledge, previous studies found that radiotherapy and chemotherapy can effectively improve the prognosis of post-operative patients with PC (24, 25). Our study found that among elderly patients with PC, those who had received chemotherapy and radiotherapy accounted for ∼55.69% and 23.03% of the total number of patients, respectively, which was consistent with the previous study. Therefore, it is necessary to establish and validate a reliable prognostic model for elderly patients with PC with better applicability.

Survival was estimated by the Kaplan–Meier method, and any differences in survival were evaluated with a stratified log-rank test. Multivariable analyses with the Cox proportional-hazards model were used to estimate the simultaneous effects of prognostic factors on survival. Interactions with prognostic factors were also examined with the Cox proportional-hazards

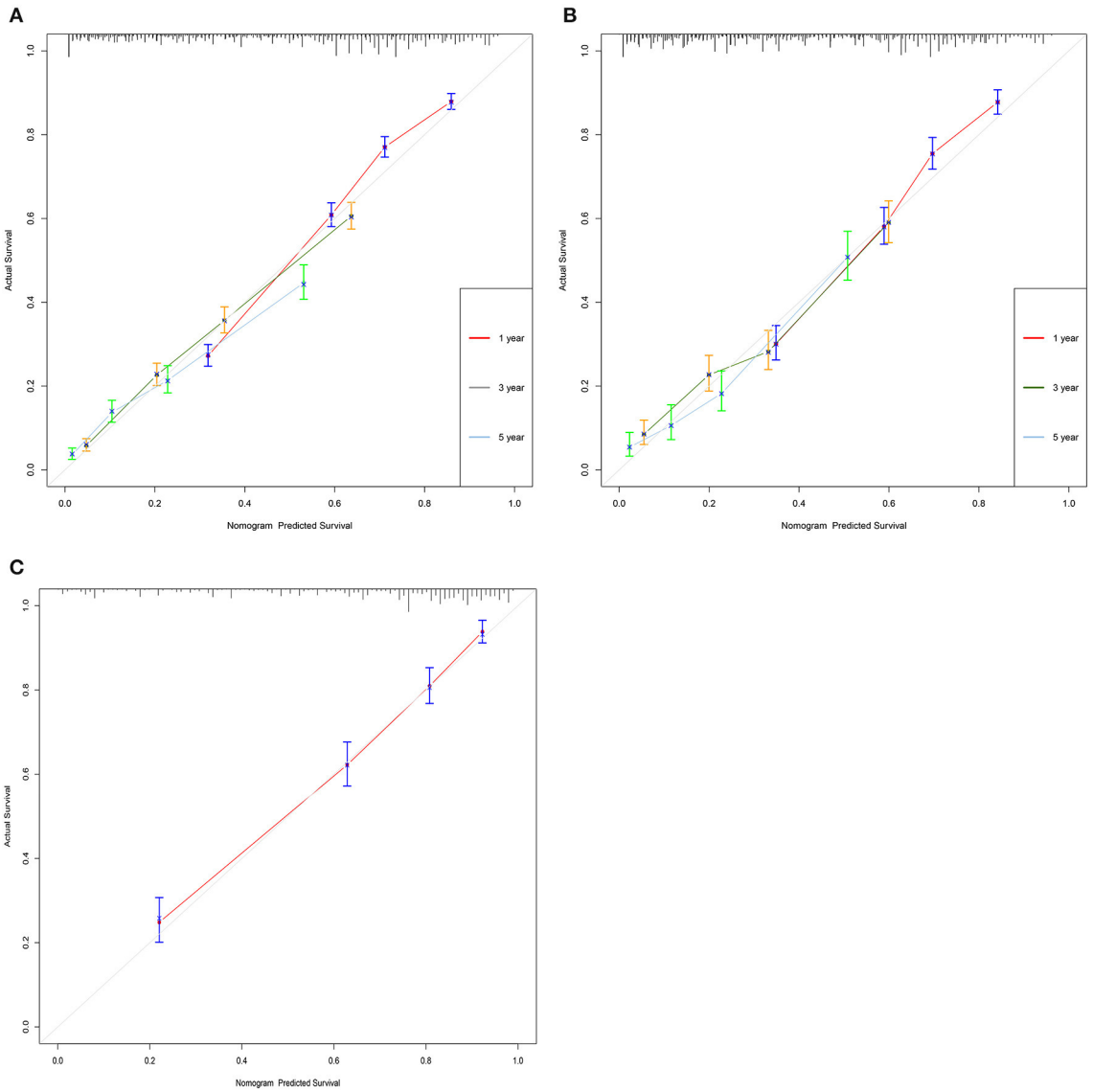


FIGURE 3 | Calibration curves of the nomogram. **(A)** For 1-, 3-, and 5-year OS in training cohort; **(B)** for 1-, 3-, and 5-year OS in validation cohort; and **(C)** for 1-, 3-, and 5-year OS in validation cohort.

model. Kaplan–Meier curves and log-rank tests study the relationship between a single variable and survival, and Kaplan–Meier and log-rank tests are only applicable to categorical variables, but not to numerical variables. The advantage of Cox proportional hazards regression analysis is that it can analyze categorical and numerical variables, and expand the scope of survival analysis from univariate to multivariate analysis (26). By univariate and multivariate Cox proportional hazard regression, several independent prognostic determinants were found to significantly affect OS, namely, age, T stage, histological grade, surgery, radiotherapy, chemotherapy, insurance, and stage of AJCC. We used these factors to construct and validate such prognostic nomograms to predict the probability of OS, and it is gratifying that the model manifested good discrimination and calibration, which meant it might act as a quantitative model to assess the probability of OS in patients with PC. At present, surgery is usually considered the optimal curative option for PC. It would always be recommended if the tumors were resectable. In our study, surgery had the greatest influence on OS in elderly patients with PC, followed by grade, AJCC stage, and T stage. As shown in this study, in most patients, resection of the masses was still the first option to achieve a longer OS with a significant p-value <0.001, indicating that extended pancreatoduodenectomy or local excision of the tumor should be considered even if the procedure is hard to conduct (27). Overall, surgical treatment has better OS than non-surgical treatment. The variance of tumor location in the pancreas can often be tackled in different ways, which determines the prognosis of

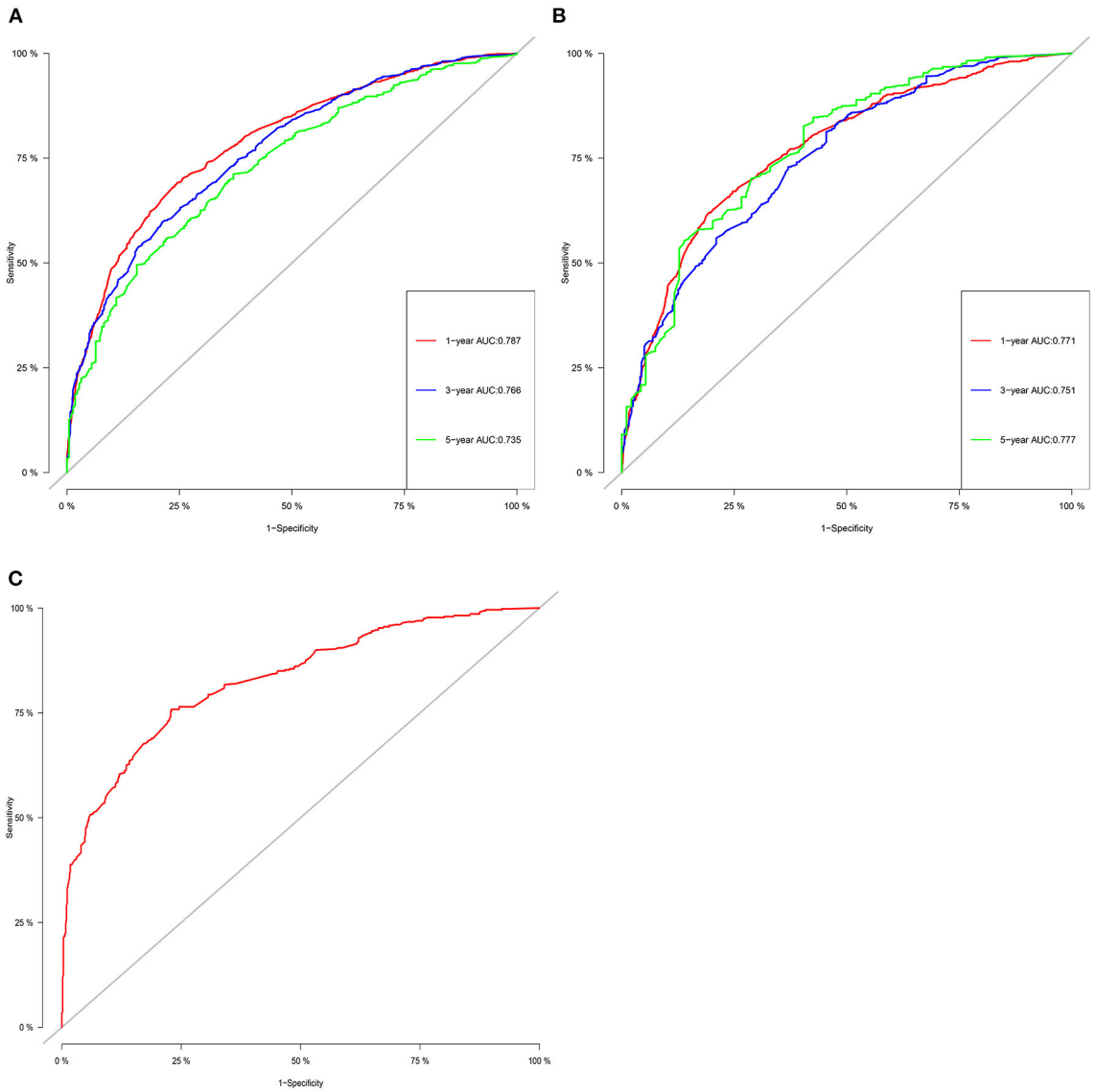


FIGURE 4 | The ROC for OS of 1-, 3- and 5-year of training cohort **(A)**, validation cohort **(B)**, and external validation cohort **(C)**.

patients. There still needs to be a larger randomized clinical research to further confirm it. Thus, the therapy selection of surgery decision was another robust indicator for prognosis in terms of the C-index.

The majority of patients with pancreatic adenocarcinoma are over 60 years of age (28, 29). In addition, it has been indicated that increasing age appears to be an important prognostic determinant that directly affects tumor-specific mortality. Further stratified survival analysis showed that patients older than 74 years had lower survival rates compared with those ages fell between 65 and 74 years old. This result resembled other studies which suggested that increasing age might helpful for the mortality of patients (14, 21). Furthermore, the previous study found that three-fourths of non-cancer-specific mortality was observed in elderly patients (13). Age-related comorbid conditions or complications may contribute to non-cancer-specific mortality in elderly patients with PC. This may be related to the decline in the function of the immune system caused by the aging of the patient, which in turn leads to the deterioration of the tumor and shortens the survival time of the patient. Therefore, it is of great significance to take non-cancer-specific mortality into account in the analysis of prognosis, especially in elderly patients.

As we knew, the TNM staging system or differentiation grade is a conventional method for doctors and researchers to evaluate cancer prognosis and select appropriate treatment strategies with respect to the survival of individual patients was still imprecise owing to it being relatively heterogeneous (30, 31). Therefore, the prognostic prediction model seems to be particularly valuable, especially for those who suffer from PC disease. In recent years, studies about the prognosis of PC

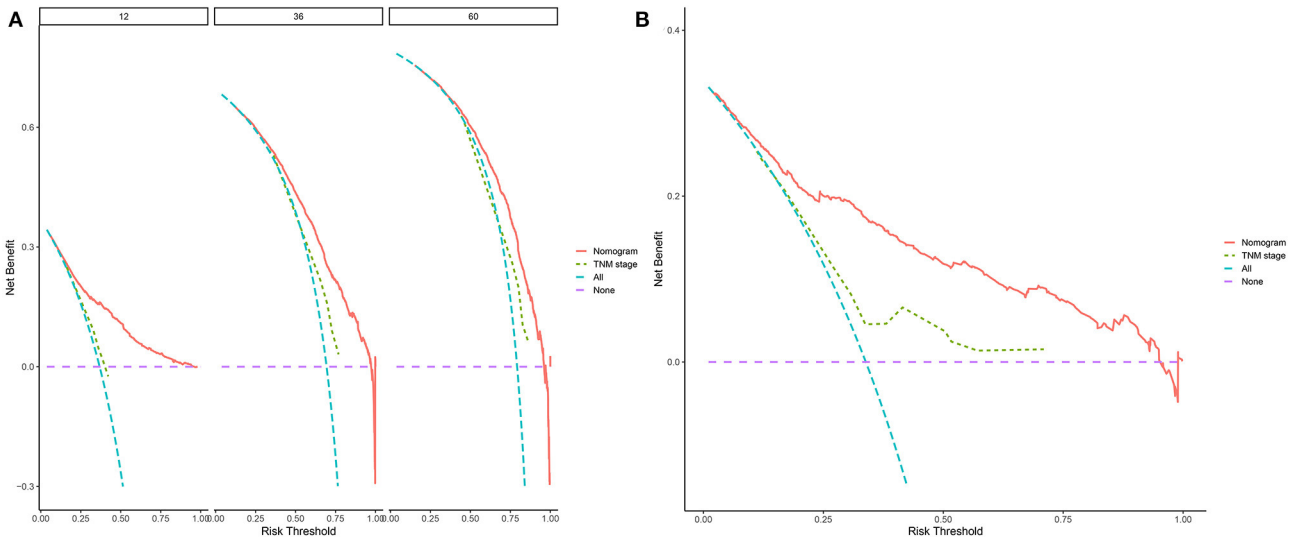


FIGURE 5 | Decision curves of the nomogram predicting OS in validation cohort **(A)** and external validation cohort **(B)**. The y-axis represents the net benefit, and the x-axis represents the threshold probability. The purple line indicates that no patients have died, and the blue line indicates that all patients have died. When the threshold probability is between 20 and 60%, the net benefit of the model exceeds all deaths or no deaths.

had been reported after the patients received surgical resection, chemotherapy, or radiotherapy (32–35). Tumor differentiation and TNM stage were commonly regarded as important factors in many cancers and some recent reports even combined these factors with other biomarkers to improve the prediction power (36–40). The performance of the nomogram was measured by the C-index and the area under the receiver operating characteristic (ROC) curve (AUC). All nomogram C-indices were >0.7, the larger the C-index, the more precise was the prognostic prediction, indicating that the model had good discriminatory power. AUC was used as a performance indicator to measure the merit of a machine learning model or a predictive model, a larger AUC (range 0.5–1.0) reflected a more accurate prediction. As indicated in this study, the ROC curves demonstrated that the nomogram showed better discrimination and the ability to provide an individualized prediction for patients. We proposed a nomogram that contains easily measurable clinical features, it is superior to the existing TNM staging system and is more conducive to patient counseling and the promotion of personalized treatment. Univariate analysis and logistic regression were performed to evaluate the impact of clinicopathologic and treatment variables on survival and find potential risk factors. Patients with advanced T stage and AJCC stage suffered from higher mortality and poorer survival rate as demonstrated by multivariate Cox regression analysis. Consistent with previous studies, carcinoma grade is a significant prognostic risk indicator (41). The nomogram model suggests that showed worse clinical outcomes when carcinoma grade shifted to poor differentiation from good differentiation.

Previous studies revealed that female, black, and unmarried patients had a worse OS (8), which was inconsistent with our study. Our study found that no significant difference existed in OS between patients with different marital statuses, gender, and race. Moreover, the multivariable analysis found that insurance has a significant impact on the OS of patients with PC. It is not difficult to imagine that medical insurance will directly affect the patient's choice of treatment, which will contribute to directly affecting the OS. Decision curve analysis (DCA) is a new calculation method that estimates the net benefits under various risk thresholds to evaluate the clinical value of the model. DCA was used to assess the clinical practicability of the model and compare it with T staging, which was a method for evaluating alternative diagnostic or prognostic tools that had advantages over others (42, 43). In this study, we used DCA to validate the accuracy and predictive ability of nomograms for elderly patients with PC, good clinical utility was indicated in the proper range. Validation of the nomogram was significant to avoid overfitting the model and determining representativeness (44). Calibration plots were used to assess the predictive accuracy of the model. In this study, the calibration curve described revealed excellent predictive performance, DCA showed superior net benefits than those of the TNM staging system, which could effectively guarantee the reliability of the constructed nomogram. Moreover, Kaplan–Meier curves showed that the OS of patients in the risk-low and risk-high groups was significantly different from each other, indicating that our model provides accurate surgical intervention and monitoring for high-risk groups, which implied good clinical application potential. Furthermore, we used data from 1,761 patients between 2016 and 2018 as external validation. The survival prediction showed that our prediction model has good accuracy and reliability. They formulated an online application for OS prediction with common and easy-accessible factors, which is convenient for clinical use.

Nevertheless, some limitations to our study should be considered as follows and also some remaining issues for the

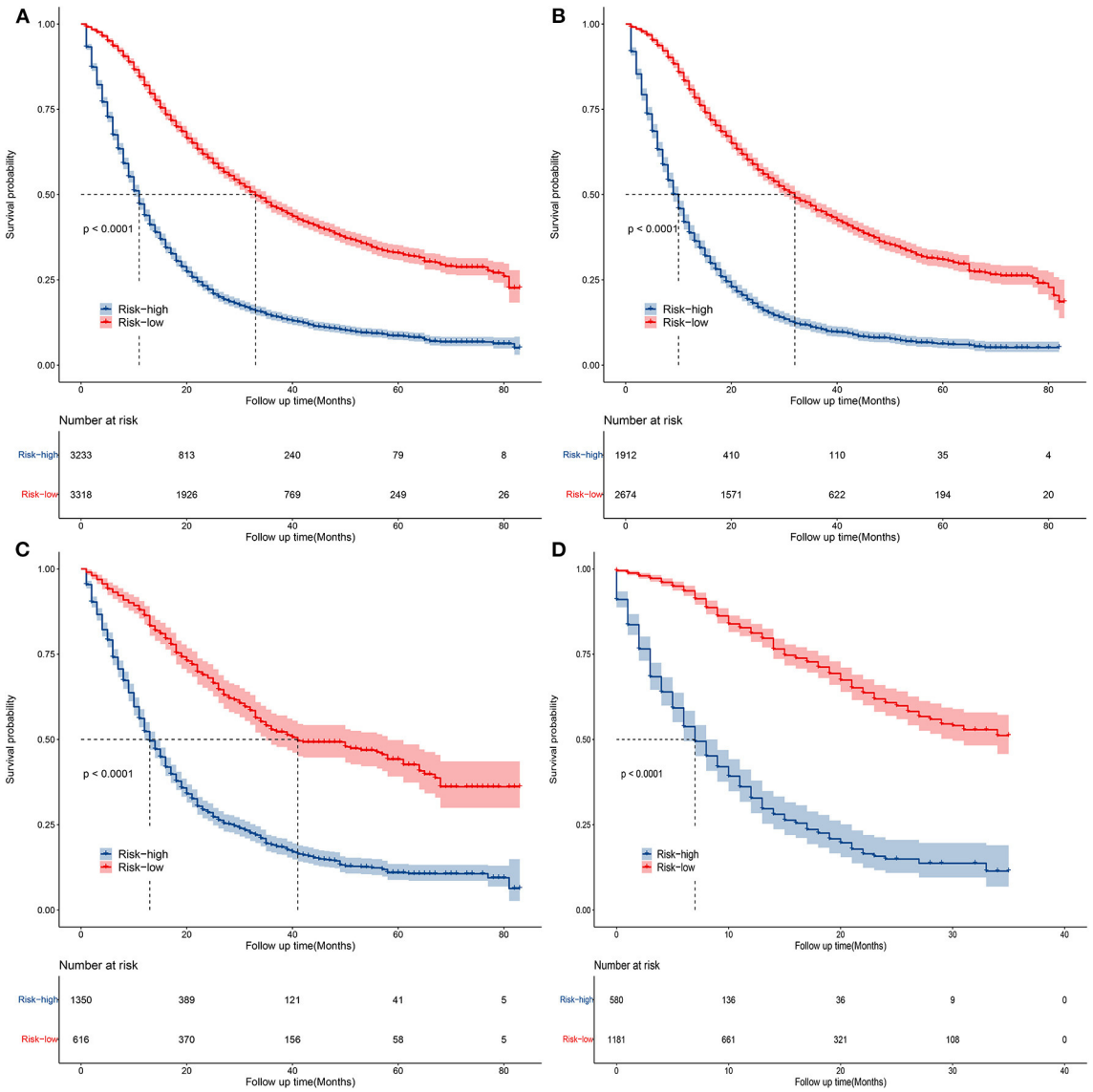


FIGURE 6 | Kaplan–Meier curves of OS for patients in the low- and high-risk groups in all cohorts **(A)**, training cohort **(B)**, validation cohort **(C)**, and external validation cohort **(D)**.

future. First, this is just a retrospective study. Next, although both the C-index and the calibration curve performed well, validation using a large other external data is still required to further evaluate the predictive accuracy and reliability of our models. Finally, other potential PC-related prognostic factors, such as the family history of PC, obesity, smoking, chronic pancreatitis, history of diabetes, pre-operative nutritional support, and vascular invasion, were not available in the SEER database.

To resolve the above-mentioned limitations, in the future, we aim at implementing the proposed model to determine its real-time applicability and scalability, and we will take other potential PC-related prognostic factors into account during the construction and validation of the model, such as the family history of PC, obesity, smoking, chronic pancreatitis, history of diabetes, and pre-operative nutritional support, lymph node metastasis, and vascular invasion. Besides, methods driven by artificial intelligence (AI) such as deep learning and reinforcement learning will be considered in the formulation of the predictive model, which can help predict specific parameters, hazards, and outcomes (45, 46). Because the blockchain is employed to ensure trust among entities, data immutability, availability, and information security, a predictive model that combines AI and blockchain technology to analyze and integrate clinical characteristics of patients will also be further studied (47, 48). AI and ML could make predictive models smarter and more interesting. Different ML concepts could be added to predictive models to induce a dynamic prediction model. Furthermore, the

feasibility of the models could also be studied by examining the computational complexity (49).

CONCLUSIONS

In conclusion, we analyzed the clinicopathological factors determining OS of elderly patients with PC using a large population-based SEER database. Cancer-specific mortality and competing risk mortality were evaluated. Furthermore, nomograms for predicting 1-, 2- and 3-year OS in elderly patients with PC were established for the first time based on a large study cohort. This constructed nomogram showed good performance and can help doctors and elderly patients with PC to more accurately and conveniently predict individual survival and formulate treatment and follow-up strategies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://seer.Cancer.gov/.

ETHICS STATEMENT

The data of this study are obtained from the SEER database. The data of patients are public and anonymous, so this study does not require ethical approval and informed consent.

AUTHOR CONTRIBUTIONS

JZ, XinL, YZ, XiaL, and SP contributed to the conception and design. XiaL, JZ, YZ, and SP collected and analyzed the data. SP and YZ drew the figures and tables. SP and XinL wrote the draft. JC, YL, CL, JQ, and XG contributed to manuscript writing and revision. All authors approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.885624/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin*. (2020) 70:7–30. doi: 10.3322/caac.21590
2. Jia X, Du P, Wu K, Xu Z, Fang J, Xu X, et al. Pancreatic cancer mortality in china: characteristics and prediction. *Pancreas*. (2018) 47:233–7. doi: 10.1097/MPA.0000000000000976
3. Ye L, Ye H, Zhou Q, Li Z, Lin Q, Tan L, et al. A retrospective cohort study of pancreatic neuroendocrine tumors at single institution over 15 years: New proposal for low- and high-grade groups, validation of a nomogram for prognosis, and novel follow-up strategy for liver metastases. *Int J Surg*. (2016) 29:108–17. doi: 10.1016/j.ijsu.2016.03.036
4. Zhao Z, Liu W. Pancreatic cancer: a review of risk factors, diagnosis, and treatment. *Technol Cancer Res Treat*. (2020) 19:1533033820962117. doi: 10.1177/1533033820962117
5. Adam MA, Thomas S, Roman SA, Hyslop T, Sosa JA. Rethinking the current american joint committee on cancer TNM staging system for medullary thyroid cancer. *JAMA Surg*. (2017) 152:869–76. doi: 10.1001/jamasurg.2017.1665
6. Choi KH, Kim BS, Oh ST, Yook JH, Kim BS. Comparison the sixth and seventh editions of the AJCC staging system for T1 gastric cancer: a long-term follow-up study of 2124 patients. *Gastric Cancer*. (2017) 20:43–8. doi: 10.1007/s10120-015-0590-0
7. Shao N, Xie C, Shi Y, Ye R, Long J, Shi H, et al. Comparison of the 7th and 8th edition of American Joint Committee on Cancer (AJCC) staging systems for breast cancer patients: a surveillance, epidemiology and end results (SEER) analysis. *Cancer Manag Res*. (2019) 11:1433–42. doi: 10.2147/CMAR.S185212
8. Shi M, Zhou B, Yang SP. Nomograms for predicting overall survival and cancer-specific survival in young patients with pancreatic cancer in the US based on the SEER database. *PeerJ*. (2020) 8:e8958. doi: 10.7717/peerj.8958
9. Bunduc Ş, Iacob R, Costache R, Stoica B, Radu C, Gheorghe C. very early onset pancreatic adenocarcinoma - clinical presentation, risk factors and therapeutic options. *Chirurgia*. (2018) 113:405–11. doi: 10.21614/chirurgia.113.3.405
10. Ansari D, Althini C, Ohlsson H, Andersson R. Early-onset pancreatic cancer: a population-based study using the SEER registry. *Langenbecks Arch Surg*. (2019) 404:565–71. doi: 10.1007/s00423-019-01810-0
11. Kang JS, Jang JY, Kwon W, Han Y, Kim SW. Clinicopathologic and survival differences in younger patients with pancreatic ductal adenocarcinoma-A propensity score-matched comparative analysis. *Pancreatology*. (2017) 17:827–32. doi: 10.1016/j.pan.2017.08.013
12. He C, Zhang Y, Cai Z, Lin X, Li S. Overall survival and cancer-specific survival in patients with surgically resected pancreatic head adenocarcinoma: a competing risk nomogram analysis. *J Cancer*. (2018) 9:3156–67. doi: 10.7150/jca.25494
13. Li J, Liu L. Overall survival in patients over 40 years old with surgically resected pancreatic carcinoma: a SEER-based nomogram analysis. *BMC Cancer*. (2019) 19:726. doi: 10.1186/s12885-019-5958-9
14. Song W, Miao DL, Chen L. Nomogram for predicting survival in patients with pancreatic cancer. *Onco Targets Ther*. (2018) 11:539–45. doi: 10.2147/OTT.S154599
15. Zhang ZY, Luo QF, Yin XW, Dai ZL, Basnet S, Ge HY. Nomograms to predict survival after colorectal cancer resection without preoperative therapy. *BMC Cancer*. (2016) 16:658. doi: 10.1186/s12885-016-2684-4
16. Liu J, Geng Q, Liu Z, Chen S, Guo J, Kong P, et al. Development and external validation of a prognostic nomogram for gastric cancer using the national cancer registry. *Oncotarget*. (2016) 7:35853–64. doi: 10.18632/oncotarget.8221
17. Li Y, Ju J, Liu X, Gao T, Wang Z, Ni Q, et al. Nomograms for predicting long-term overall survival and cancer-specific survival in patients with major salivary gland cancer: a population-based study. *Oncotarget*. (2017) 8:24469–82. doi: 10.18632/oncotarget.14905
18. He T, Chen T, Liu X, Zhang B, Yue S, Cao J, et al. A web-based prediction model for cancer-specific survival of elderly patients with early hepatocellular carcinoma: a study based on SEER database. *Front Public Health*. (2022) 9:789026. doi: 10.3389/fpubh.2021.789026
19. Brateanu A, Yu C, Kattan MW, Olender J, Nielsen C. A nomogram to predict the probability of passing the American Board of Internal Medicine examination. *Med Educ Online*. (2012) 17:18810. doi: 10.3402/meo.v17i0.18810
20. Cao J, Yuan P, Wang L, Wang Y, Ma H, Yuan X, et al. clinical nomogram for predicting survival of esophageal cancer patients after esophagectomy. *Sci Rep*. (2016) 6:26684. doi: 10.1038/srep26684
21. Pu N, Lv Y, Zhao G, Lee W, Nuerxiati A, Wang D, et al. Survival prediction in pancreatic cancer patients with no distant metastasis: a large-scale population-based estimate. *Future Oncol*. (2018) 14:165–75. doi: 10.2217/fon-2017-0380
22. Ntala C, Debernardi S, Feakins RM, Crnogorac-Jurcevic T. Demographic, clinical, and pathological features of early onset pancreatic cancer patients. *BMC Gastroenterol*. (2018) 18:139. doi: 10.1186/s12876-018-0866-z
23. Beeghly-Fadiel A, Luu HN, Du L, Shi C, McGavic DP, Parikh AA, et al. Early onset pancreatic malignancies: clinical characteristics and survival associations. *Int J Cancer*. (2016) 139:2169–77. doi: 10.1002/ijc.30273

24. Springfeld C, Jäger D, Büchler MW, Strobel O, Hackert T, Palmer DH, et al. Chemotherapy for pancreatic cancer. *Presse Med*. (2019) 48:e159–74. doi: 10.1016/j.lpm.2019.02.025
25. Stessin AM, Meyer JE, Sherr DL. Neoadjuvant radiation is associated with improved survival in patients with resectable pancreatic cancer: an analysis of data from the surveillance, epidemiology, and end results (SEER) registry. *Int J Radiat Oncol Biol Phys*. (2008) 72:1128–33. doi: 10.1016/j.ijrobp.2008.02.065
26. Cox, David R. Regression models and life-tables. *J R Stat Soc Ser B*. (1972) 34:187–202. doi: 10.1111/j.2517-6161.1972.tb00899.x
27. Choti MA, Dixon E, Tyler D. Pretreatment assessment of resectable and borderline resectable pancreatic cancer: expert consensus statement by Callery et al. *Ann Surg Oncol*. (2009) 16:1734–5. doi: 10.1245/s10434-009-0411-y
28. Hamada T, Nakai Y, Yasunaga H, Isayama H, Matsui H, Takahara N, et al. Prognostic nomogram for nonresectable pancreatic cancer treated with gemcitabine-based chemotherapy. *Br J Cancer*. (2014) 110:1943–9. doi: 10.1038/bjc.2014.131
29. Eubanks A, Pepe J, Veldhuis P, de la Fuente SG. Age as a prognostic indicator for adjuvant therapy in patients who underwent pancreatic resections for cancer. *J Geriatr Oncol*. (2018) 9:362–6. doi: 10.1016/j.jgo.2017.12.004
30. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med*. (2015) 373:1803–13. doi: 10.1056/NEJMoa1510665
31. Diwakarla C, Hannan K, Hein N, Yip D. Advanced pancreatic ductal adenocarcinoma - complexities of treatment and emerging therapeutic options. *World J Gastroenterol*. (2017) 23:2276–85. doi: 10.3748/wjg.v23.i13.2276
32. An X, Ding PR, Li YH, Wang FH, Shi YX, Wang ZQ, et al. Elevated neutrophil to lymphocyte ratio predicts survival in advanced pancreatic cancer. *Biomarkers*. (2010) 15:516–22. doi: 10.3109/1354750X.2010.491557
33. Xue P, Kanai M, Mori Y, Nishimura T, Uza N, Kodama Y, et al. Neutrophil-to-lymphocyte ratio for predicting palliative chemotherapy outcomes in advanced pancreatic cancer patients. *Cancer Med*. (2014) 3:406–15. doi: 10.1002/cam4.204
34. Ji F, Fu SJ, Guo ZY, Pang H, Ju WQ, Wang DP, et al. Prognostic value of combined preoperative lactate dehydrogenase and alkaline phosphatase levels in patients with resectable pancreatic ductal adenocarcinoma. *Medicine*. (2016) 95:e4065. doi: 10.1097/MD.0000000000004065
35. Shirai Y, Shiba H, Haruki K, Horiuchi T, Saito N, Fujiwara Y, et al. Preoperative platelet-to-albumin ratio predicts prognosis of patients with pancreatic ductal adenocarcinoma after pancreatic resection. *Anticancer Res*. (2017) 37:787–93. doi: 10.21873/anticanres.11378
36. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, et al. Development and validation of a nomogram for predicting survival in patients with resected non-small-cell lung cancer. *J Clin Oncol*. (2015) 33:861–9. doi: 10.1200/JCO.2014.56.6661
37. He X, Guo S, Chen D, Yang G, Chen X, Zhang Y, et al. Preoperative albumin to globulin ratio (AGR) as prognostic factor in renal cell carcinoma. *J Cancer*. (2017) 8:258–65. doi: 10.7150/jca.16525
38. Jing CY, Fu YP, Shen HJ, Zheng SS, Lin JJ, Yi Y, et al. Albumin to gamma-glutamyltransferase ratio as a prognostic indicator in intrahepatic cholangiocarcinoma after curative resection. *Oncotarget*. (2017) 8:13293–303. doi: 10.18632/oncotarget.14530
39. Chan A, Prassas I, Dimitromanolakis A, Brand RE, Serra S, Diamandis EP, et al. Validation of biomarkers that complement CA19.9 in detecting early pancreatic cancer. *Clin Cancer Res*. (2014) 20:5787–95. doi: 10.1158/1078-0432.CCR-14-0289
40. Stotz M, Szkandera J, Stojakovic T, Seidel J, Samonigg H, Kornprat P, et al. The lymphocyte to monocyte ratio in peripheral blood represents a novel prognostic marker in patients with pancreatic cancer. *Clin Chem Lab Med*. (2015) 53:499–506. doi: 10.1515/cclm-2014-0447
41. Leibovich BC, Lohse CM, Crispen PL, Boorjian SA, Thompson RH, Blute ML, et al. Histological subtype is an independent predictor of outcome for patients with renal cell carcinoma. *J Urol*. (2010) 183:1309–15. doi: 10.1016/j.juro.2009.12.035
42. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. (2006) 26:565–74. doi: 10.1177/0272989X06295361
43. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. (2008) 8:53. doi: 10.1186/1472-6947-8-53
44. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791
45. Kumar RL, Khan F, Din S, Band SS, Mosavi A, Ibeke E. Recurrent neural network and reinforcement learning model for COVID-19 prediction. *Front Public Health*. (2021) 9:744100. doi: 10.3389/fpubh.2021.744100
46. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT, Alinejad-Rokny H. A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues. *J Biomed Inform*. (2021) 113:103627. doi: 10.1016/j.jbi.2020.103627
47. Samuel O, Omojo AB, Onuja AM, Sunday Y, Tiwari P, Gupta D, et al. IoMT: a COVID-19 healthcare system driven by federated learning and blockchain. *IEEE J Biomed Health Inform*. (2022) PP:10.1109/JBHI.2022.3143576. doi: 10.1109/JBHI.2022.3143576
48. Mani V, Kavitha C, Band SS, Mosavi A, Hollins P, Palanisamy S. A recommendation system based on AI for storing block data in the electronic health repository. *Front Public Health*. (2022) 9:831404. doi: 10.3389/fpubh.2021.831404
49. Islam MJ, Rahman A, Kabir S, Karim MR, Acharjee UK, Nasir MK, et al. Blockchain-SDN based energy optimized and distributed secure architecture for IoTs in smart cities. *Preprints*. (2020) 11:0552. doi: 10.20944/preprints202011.0552.v1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhong, Liao, Peng, Cao, Liu, Liu, Qiu, Guan, Zhang, Liu and Peng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 25 May 2022
doi: 10.3389/fpubh.2022.916407

Analysis of the Impacts on the Psychological Changes of Chinese Returning College Students After the Outbreak of the 2019 Coronavirus Disease

Yingying Xue[1,2], *Kwak Han Pyong*[2], *Sae Sook Oh*[2], *Yingying Tao*[3] *and Taofeng Liu*[4,5*]

[1] Institute of Physical Education, Yangzhou University, Yangzhou, China, [2] Sports Science Department, Kyonggi University, Suwon, South Korea, [3] Department of College Physical Education, Communication University of Zhejiang, Hangzhou, China, [4] Zhengzhou University Physical Education Institute (Main Campus), Zhengzhou, China, [5] Department of Physical Education, Sangmyung University, Seoul, South Korea

OPEN ACCESS

Edited by:
Ali Kashif Bashir, Manchester Metropolitan University, United Kingdom

Reviewed by:
Lin Haifeng, McMaster University, Canada
Wnieee Wnieee, South-Central University for Nationalities, China
Hao Chen, Hunan University, China

***Correspondence:**
Taofeng Liu
202033084@sangmyung.kr

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 April 2022*
Accepted: *28 April 2022*
Published: *25 May 2022*

Citation:
Xue Y, Pyong KH, Oh SS, Tao Y and Liu T (2022) Analysis of the Impacts on the Psychological Changes of Chinese Returning College Students After the Outbreak of the 2019 Coronavirus Disease. Front. Public Health 10:916407. doi: 10.3389/fpubh.2022.916407

This work aims to analyze the impacts on the psychological changes of Chinese returning college students after the outbreak of the 2019 coronavirus disease (COVID-19). A questionnaire survey is used to take 1,482 college students who returned to school after the epidemic as the research objects. The Chinese college students' knowledge of the epidemic, alienation in physical education class, school happiness, and expectations for a healthy life in the future are investigated and analyzed. The research results manifest that Chinese returning college students have relatively poor awareness of COVID-19, and the overall degree of alienation in physical education classes after the epidemic is low, with an average score of 3.55 ± 1.018. The overall level of school happiness is high, with an average score of 4.94 ± 0.883; the overall level of expectation for a healthy life in the future is high, with an average score of 3.50 ± 0.840. It denotes that the epidemic has a great psychological impact on returning college students, and it is necessary to strengthen mental health education for college students after COVID-19. It provides a sustainable theoretical reference for the formulation of psychological intervention measures for returning college students.

Keywords: COVID-19, college students, alienation, school happiness, psychological changes

INTRODUCTION

A major infectious disease broke out in China at the end of 2019, that is, pneumonia caused by a new type of coronavirus, which is called "Corona Virus Disease 2019" by the World Health Organization (WHO) (1, 2). The outbreak of Corona Virus Disease 2019 occurred in December 2019, mainly in Wuhan City, Hubei Province, China. The cause of the disease was unknown. Epidemiological investigations found that they had all been to the Huanan Seafood Market, and were finally diagnosed as acute respiratory disease caused by new coronavirus infection (3). Comrade Tedros of the WHO announced in Geneva, Switzerland on February 11, 2020, that the disease was named as "COVID-19" (4). Meanwhile, the National Health Commission (NHC) issued a notice on February 22 to change its English name to "COVID-19." After that, WHO declared the

current COVID-19 outbreak a global pandemic on March 11 (5, 6). COVID-19 is an infectious disease of the respiratory tract caused by coronavirus type 2 infection, which can induce severe acute respiratory syndrome (7). Since the onset of the disease, all provinces, municipalities, and autonomous regions in China have taken strict prevention and control measures to avoid travel and crowd gathering as much as possible, and have achieved good results. According to the existing case data, the main manifestations of COVID-19 include fever, dry cough, and fatigue. A small number of patients will have symptoms of the upper respiratory and digestive systems, such as nasal congestion, runny nose, and diarrhea. Severely ill patients usually suffer symptoms of dyspnea after 7 days of onset, can be complicated by Acute Respiratory Distress Syndrome (ARDS) in a short period of time, and then develop septic shock and metabolic acidosis (refractory), multiple organ failure, etc. (8–10). COVID-19 is a Class B infectious disease managed at the national level, and a public health emergency. Since the establishment of the People's Republic of China, it has spread the fastest, and it is the most widespread and the most difficult to prevent and control infection among major public health events (11). College students are in an active period of physical and mental development. Faced with the dual pressures of the epidemic and their studies, college students are more likely to have negative emotions, develop bad living and exercise habits, and decline in physical fitness.

So far, measures for the prevention and control of the COVID-19 epidemic taken by Chinese government have developed from the initial emergency state to the current normalization work. Full-scale implementation began, and in the meantime, the resumption of normal work and production has been actively underway (12). The Guidelines for the Prevention and Control of Novel Coronavirus Pneumonia in Colleges and Universities proposed in March 2020 stated that the huge pressure caused by the new coronary pneumonia can easily induce psychological stress such as anxiety, depression, and hypochondriasis. During the period of the epidemic, students were isolated and managed at home. The government and schools encouraged and supervised students to exercise through online physical education teaching and assignment of exercise homework. However, home exercise mainly relies on students' conscious self-exercise, and schools lack effective monitoring methods (13, 14). This has a great impact on college students' self-efficacy, life behavior, mental health, learning adaptability, social responsibility and values (15, 16). With the continuous progress and development of information technology, the research methods in psychology have also undergone great changes since the 1940s (17). In the 1970s, some experts tried to use computers to carry out experiments in psychology, which led to the first revolution in the psychology. In the 1990s, the birth of the online psychological questionnaire was the second revolution in psychology due to the advancement of network technology. At this stage, experts and scholars in the psychology established virtual psychological laboratories on the Internet. The advent of the era of big data has prompted the beginning of the third revolution in psychology. The samples of big data are featured with large amount of data, various types, fast traffic, and great potential significance. They have gradually become valuable resources for researchers to observe the characteristics and laws of human psychological behaviors (18–21).

Compared with previous studies, psychological research based on big data technology has great advantages. This shows that psychological research in the past has focused more on the explanation of a certain psychological response. Big data technology can predict and intervene certain psychological reactions on the basis of interpretation (22). Furthermore, the research method based on big data technology can conduct an overall analysis of all data, avoid the sampling error of previous research methods from the sample to the overall promotion, reduce the experimental error caused by improper control of experimental conditions, and greatly improve the efficiency and scale of data processing. In terms of time validity, real-time collection and analysis of psychological states avoid the deviation of analysis results caused by the lag of data collection in traditional research methods (23, 24). The innovation is the use of regression analysis to examine the status of COVID-19-related knowledge among Chinese returning college students after the COVID-19 pandemic, as well as changes in their alienation in physical education classes, school happiness, and expectations for a healthy life in the future. This study aims to provide a sustainable theoretical reference for the formulation of psychological intervention measures for returning college students. The research framework is to first introduce the research objects, observation indicators and survey methods of the questionnaire; then regression analysis is used to analyze the changes in Chinese returning college students' alienation in physical education classes, school happiness, and expectations for a healthy life in the future after the COVID-19 pandemic.

MATERIALS AND METHODS

Research Objects

The random sampling is adopted to investigate the college students who returned to school after the COVID-19 pandemic in different provinces and cities in China. From June to October 2020, a questionnaire survey was conducted on 1,482 college students who were approved to return to school under the principle of informed consent. The included college students have to meet the following criteria: full-time college students returning to school after the epidemic (undergraduate, master, and above) and returning college students who are informed with this experiment and cooperate with the investigation. If any student meets the below items, it has to be excluded: students who have left or dropped out of school; and students who are on-the-job, self-taught adults, or receiving correspondence education. All respondents have fully understood the situation and signed the informed consent forms, and the Medical Ethics Committee of our hospital has also been informed and agreed to implement it.

Observation Indicators

The general information of included students is investigated, including gender, age, grade, major, and place of origin. Cognition of COVID-19 includes what type of infectious disease is the COVID-19 and its route of transmission, the

main symptoms, the method of transmission, and the personal preventive measures. The situation related to the COVID-19 epidemic includes what information about the epidemic situation is the most concerned about, how much time it takes to pay attention to the epidemic situation every day, how to obtain relevant information about the epidemic situation, and the attention in information about the epidemic situation reported by the media. College student alienation scale (25): there are 52 questions in total, the higher the score, the stronger the alienation. The school happiness scale for college students (26): there are 36 questions, the higher the score, the higher the level of happiness. The expected measure of future healthy life (27): there are 16 questions in total, the higher the score, the higher the expectation.

Investigation Methods

The questionnaire survey is carried out by scanning the questionnaire star QR code with a smartphone. After the consents of the participants are obtained, the mobile phone WeChat is adopted to scan the code to open the general information about returning college students after the COVID-19 pandemic, COVID-19 awareness, epidemic-related situations, alienation in physical education classes, school happiness, and future life expectations. The researcher writes the instruction language uniformly, follows the principle of free and informed consent, and explains the purpose of the questionnaire survey and the confidentiality principle in detail to the participants. In the context of the questionnaire star operation, effective permission settings are carried out, that is, an IP address can only answer the questionnaire online once, and the filling time is not $<$5 min.

Linear Regression Analysis

Linear regression refers to a regression analysis that simulates the relationship between one or more independent variables and dependent variables through the least squares function obtained from the linear regression equation (28). This least squares function is a linear combination of model parameters and one or more regression coefficients. There are also such relationships in our daily life, such as the irrelevant relationship between random numbers, the uncertain relationship between individual height/weight, and the definite relationship between the area/radius of a circle. Regression analysis is one of the mathematical methods that is often used to study the existence or absence of correlations between variables that are not completely determined. When only one independent variable is used for regression analysis, it is called univariate linear regression analysis; when there are multiple variables, it is called multiple linear regression analysis.

I. Univariate linear regression analysis

Univariate linear regression analysis, also known as Simple Linear Regression (SLR), is the simplest regression analysis, and there are regression models that are widely used in various fields of life. Its regression equation is given as follows:

$$A{=}\alpha + \beta B + \varepsilon \tag{1}$$

In general, the least squares method is used to evaluate the minimum errors α and β from the sample (a_i, b_i) (i = 1,2,…,n), ε represents the residual, and the objective is to minimize the residual sum of squares:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(a_i - \alpha - \beta b_i\right)^2 \tag{2}$$

The extreme value can be obtained by using the differential method. In equation (2), the first-order partial differential is made with α and β, respectively, and it is equal to 0 to obtain $\widehat{\alpha}$ and $\widehat{\beta}$:

$$\widehat{\alpha} = \bar{a} - \bar{b}\widehat{\beta} \tag{3}$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} \left(b_i - \bar{b}\right)(a_i - \bar{a})}{\sum_{i=1}^{n} \left(b_i - \bar{b}\right)^2} \tag{4}$$

II. Multiple linear regression analysis

The equation for the multiple linear regression prediction model is expressed as follows:

$$\widehat{A}_t = \alpha + \beta_1 b_1 + \beta_2 b_2 + \beta_3 b_3 + ... + \beta_n b_n \tag{5}$$

b_1-b_n represent independent variables; α, β_1-β_n express the parameters of the linear regression equation. The binary linear regression analysis model with only two independent variables is the general form of the multiple linear regression model, which is a method for analyzing the correlation between two independent variables and one dependent variable. Its regression equation is given in equation below:

$$\widehat{A}_t = \alpha + \beta_1 b_1 + \beta_2 b_2 \tag{6}$$

In the above equation, $\widehat{A}_t$ represents the dependent variable, b_1 and b_2 represents the independent variables; α, β_1, and β_2 indicate the parameters of the linear regression equation. α, β1, and β2 can be obtained by solving below equation:

$$\begin{cases} \sum a = n\alpha + \beta_1 \sum b_1 + \beta_2 \sum b_2 \\ \sum b_1 a = \alpha \sum b_1 + \beta_1 \sum b_1^2 + \beta_2 \sum b_1 b_2 \\ \sum b_2 a = \alpha \sum b_2 + \beta_1 \sum b_1 b_2 + \beta_2 \sum b_2^2 \end{cases} \tag{7}$$

Ensemble Learning Method

There is a famous theory of "no free lunch" in machine learning, which states that no algorithm can maintain its accuracy all the time. Therefore, the concept of ensemble learning appears (29). The main idea of ensemble learning is to use a certain strategy to aggregate multiple base classifiers to obtain higher accuracy. Different base classifiers can use different algorithms, parameters, features, training data sets, etc. The basic idea of Bagging, Boosting, and Stacking is to integrate three strategies.

Among them, the basic idea of the Stacking strategy (30) is to train a variety of base classifiers $\{D_1, D_2, ..., D_N\}$. In addition, the Stacking strategy uses the output results $\{q_{i1}, q_{i2}, ..., q_{iN}, a_i\}$ of these base classifiers as input data, retrains to obtain a new model, and obtains the final result. output result.

Boosting tree is a boosting method based on the decision tree function. When it faces the classification task, this decision tree represents a binary decision tree; In the regression problem, the decision tree represents the regression tree, and the boosting tree can be regarded as an additive model of the decision tree:

$$f_N(b) = \sum_{n=1}^{N} T(b; \Theta_n) \tag{8}$$

In equation (8), Θ_n refers to the model parameters, $T(b; \Theta_n)$ denotes the decision tree model, and N is the number of trees.

When binary classification is required, boosting tree is a special case of binary decision trees specified by the base classifier of the AdaBoost algorithm.

Gradient Boosting Decision Tree (GBDT) is an optimization that can be used to solve the optimization problem when the loss function of the boosting tree is a general function (i.e., a non-square loss function or a non-exponential loss function). The GBDT algorithm flow is described as follows. The value of the negative gradient of the loss function in this model plays a key role in the gradient boosting algorithm, which is an algorithm that is very similar to the steepest descent method:

$$-\left[\frac{\partial L(a_i, F(b_i))}{\partial F(b_i)}\right]_{F(b)=F_{k-1}(b)} \tag{9}$$

When faced with a binary classification, the loss function that can be defined is as follows:

$$L(a, F(b)) = \log(1 + \exp(-2aF)), a \in \{+1, -1\} \tag{10}$$

The process of GBDT is as follows:

Input: training dataset $D = \{(b_1, a_1), (b_2, a_2), ..., (b_N, a_N)\}$, a function that is capable of micro-loss is $L(a, F(b))$, and the number of iterations is K.

Output: the GBDT model $F_K(b)$.

The model is initialized as follows:

$$F_0(b) = \arg_\gamma \min \sum_{i=1}^{n} L(a_i, \gamma) \tag{11}$$

k = 1,2,...,K is cycled:

(a) The desired pseudo-residuals are calculated with equation (12) below:

$$r_{ik} = -\left[\frac{\partial L(a_i, F(b_i))}{\partial F(b_i)}\right]_{F(b)=F_{k-1}(b)}, for i = 1, ..., n. \tag{12}$$

(b) The pseudo-residual is fitted to the base classifier $g_k(b)$, that is, the training dataset $\{(b_i, r_{ik})\}_{i=1}^{n}$ is adopted to train the base classifier.

(c) The multiplier γ_k can be calculated by solving a one-dimensional optimization:

$$\gamma_k = \arg_\gamma \min \sum_{i=1}^{n} L(a_i, F_{k-1}(b_i) + \gamma g_k(b_i)) \tag{13}$$

(d) The model is updated:

$$F_k(b) = F_{k-1}(b) + \gamma_k g_k(b) \tag{14}$$

Statistical Analysis

The SPSS 19.0 statistical analysis software package is adopted to enter and analyze the data in this experiment, *t*-test or variance analysis is applied to judge the differences between variables, and the multiple linear regression is selected to analyze the changes on dimensions of alienation and school happiness, expectations for a healthy life in the future. When $P < 0.05$, the difference is statistically significant.

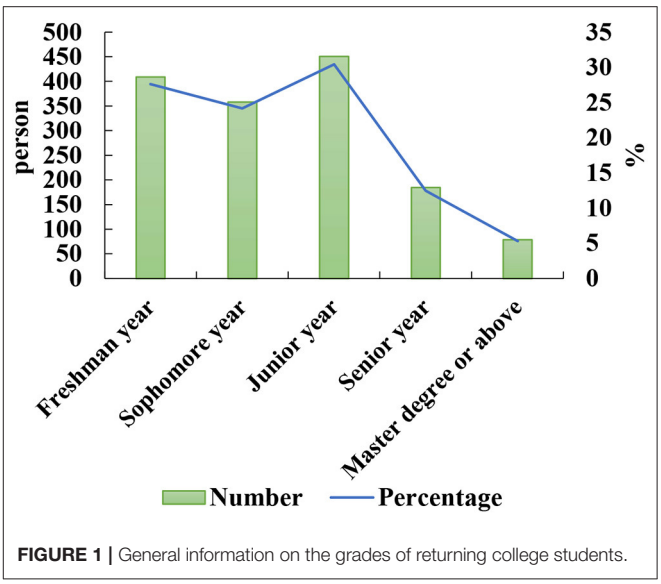


FIGURE 1 | General information on the grades of returning college students.

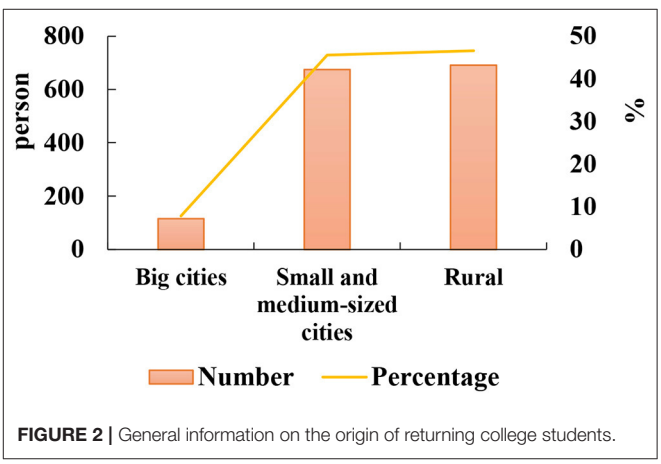


FIGURE 2 | General information on the origin of returning college students.

RESULTS

General Data

After the COVID-19 pandemic, the general information of Chinese returning college students (n = 1,482) is collected, as shown in **Figures 1**, **2**.

Among them, the numbers of males and females are 519 and 963, respectively, aged between 18 and 30 years old, with liberal arts students accounting for 41.26% and science students accounting for 58.74%. In **Figures 1**, **2**, the proportion of freshman students is 27.6%, junior students is 30.43%, and graduate students and above is 5.33%. Among returning college students, 7.83% came from big cities, 45.55% from small and medium cities, and 46.62% from rural areas.

Cognition of COVID-19

After the COVID-19 pandemic, we investigated Chinese returning college students' cognition of the disease and found that 96.7% knew that during the COVID-19 epidemic, the protective measures that people need to do include 6 aspects; and only 10.78% knew that the main manifestations of COVID-19 infection included 3 symptoms. **Figure 3** showed the details.

In the **Figure 3** above, A represents the protective measures that people need to do during the COVID-19 epidemic: washing hands frequently, refusing to eat wild animals, wearing disposable medical masks when entering and leaving public places, trying to avoid entering and leaving crowded places, staying within one meter of people around, opening the window regularly to ventilate, and reduce going out. B represents the main mode of transmission of COVID-19: close contact between people and respiratory droplet transmission. C stands for low case fatality rate of COVID-19; D represents the category B infectious disease of COVID-19 according to the prevention and control management of the category A infectious disease; E represents that if you know that in the absence of effective protective measures, if you have face-to-face communication with infected people, share meals, take public transportation, eat wild animals, or accept express delivery from areas with high epidemic incidence, it will increase the risk of epidemic transmission. F stands for the main clinical manifestations of COVID-19 infection, including fever, dry cough, and fatigue.

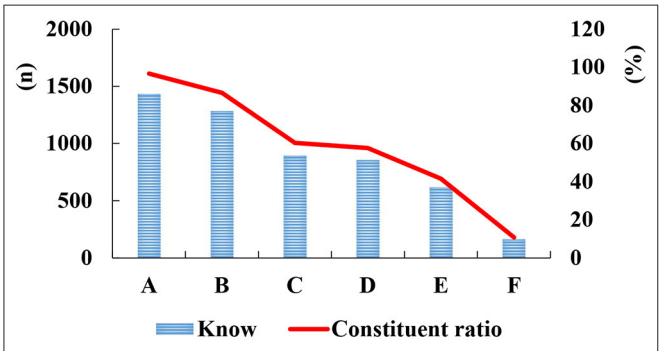


FIGURE 3 | The cognition of Chinese returning college students on the disease after the COVID-19 pandemic.

Situation of COVID-19

After the COVID-19 pandemic, we investigated the three main ways that Chinese returning college students obtain information related to COVID-19: web browsing (91.70%, 1,359/1,482), social tools such as QQ or WeChat (77.68%, 1,151/1,482), television / radio / billboard / billboard (74.8%, 1,108/1,482). The average time spent paying attention to the epidemic every day is 24.50 ± 22.6 min. The most concerned epidemic information includes: epidemic situation (58.86%, 872/1,482), research progress (22.87%, 339/1,482), and prevention and control knowledge (21.04%, 312/1,482). Concern on media reports on epidemic-related information is explained as: very concerned (73.14%, 1,084/1,482), generally concerned (25.48%, 378/1,482), and rarely concerned (4.05%, 600/1,482). The details are given in **Figure 4**.

In the above figure, **Figure 4A** shows the three main ways Chinese returning college students obtain information about COVID-19; **Figure 4B** shows the epidemic information they are

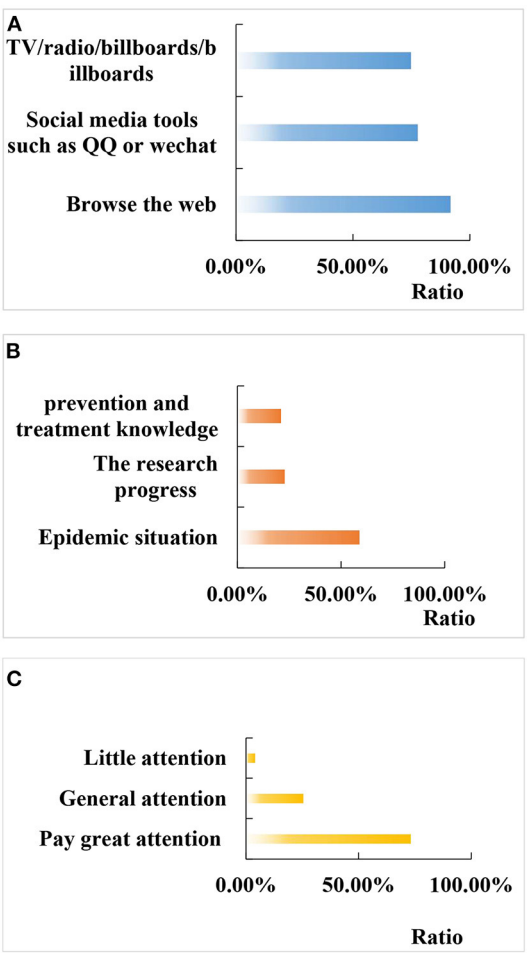


FIGURE 4 | The situation related to the COVID-19. **(A)** Main access way; **(B)** most concerned epidemic information; **(C)** attention to the information related to the epidemic situation reported by the media.

TABLE 1 | Comparison of alienation and school happiness in demographic data.

Alienation and school happiness	Difference in demographic data
Alienation	The score of boys (4.52 points in average) is higher than that of girls (4.31 points).
	The average score of Freshmen (3.98 points) is lower than the score of other grades (4.39 points).
	The average score of non-only children (4.45 points) is higher than that of only children (4.26 points).
	The average score of students living in rural areas (4.38 points) is higher than other places (4.07 points in average).
School happiness	The score of upper class students (average 5.82 points) is higher compared with that of lower class students (5.13 points in average).
	The score of students living in urban areas (with the average of 5.74 points) is slightly higher in contrast to that in rural and urban areas (5.25 points).

most concerned about; and **Figure 4C** shows their attention to media reports on epidemic-related information.

Differences in Alienation and School Happiness in Demographic Data

After the COVID-19 pandemic, the alienation in physical education class and school happiness of Chinese returning college students are compared in the term of demographic data. The results are shown in **Table 1**. It demonstrates that that among returning college students, the average alienation in physical education class is higher for boys than for girls. The average of freshmen is lower than that of other grades, the average of non-only children is higher than that of only-children, and the average score of students living in rural areas is higher than in other places. In terms of school happiness, among returning college students, the score of upper class students is higher compared with that of lower class students, and the score of students living in urban areas is slightly higher in contrast to that in rural and urban areas.

Scores of Three Scales for Returning College Students

After the COVID-19 pandemic, the overall degree of alienation in the physical education class of returning college students in China is relatively low, with an average score of 3.55 ± 1.018; the overall level of school happiness is relatively high, with an average score of 4.94 ± 0.883; and the overall level of expectations for a healthy life in the future is high (3.50 ± 0.840 in average). The specific results were shown in **Figure 5**.

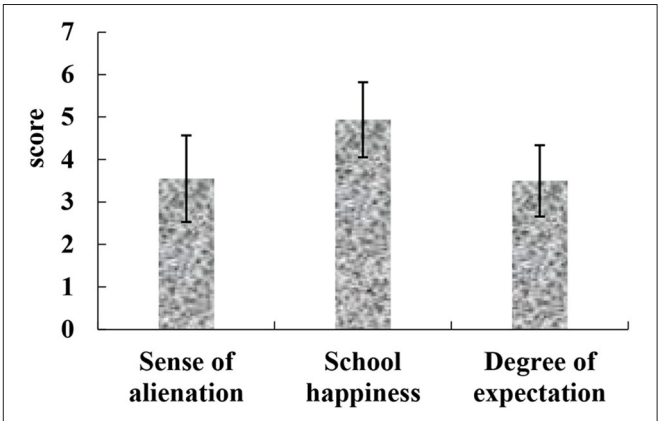


FIGURE 5 | The situation of the three scales of returning college students.

Regression Analysis

After the COVID-19 pandemic, the total score of college students returning to school in the alienation in physical education class is undertaken as the dependent variable and the school happiness and expectations for a healthy life in the future are taken as independent variables for multiple linear regression analysis. Stepwise regression method is used to screen out the optimal combined variable mode. The independent variables included in the model are shown in **Figure 6**.

In the above figure, A represents friendship happiness, B represents academic happiness, C represents freedom happiness, D represents environmental happiness, and E represents expectations for a healthy life in the future.

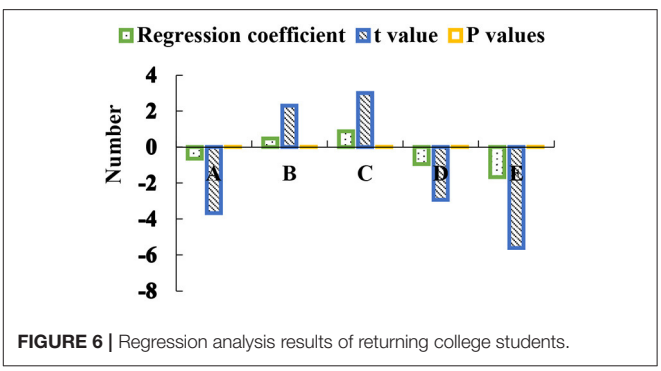
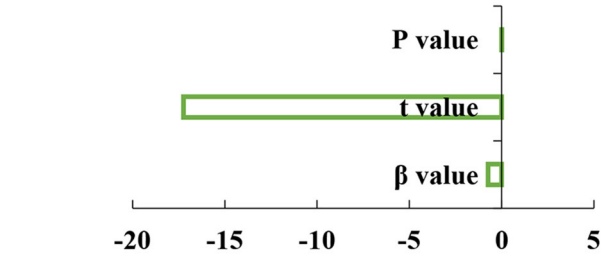


FIGURE 6 | Regression analysis results of returning college students.

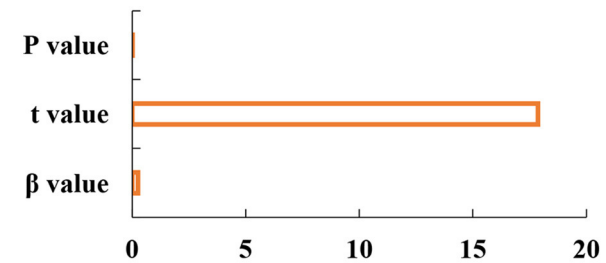
Analysis of Mediation Effect

The test procedure of the mediation effect should be followed. The first step: school happiness is selected as the independent variable and alienation is undertaken as the dependent variable to start the regression analysis with the equation of y=-0.723x ($P < 0.05$). The second step: the school happiness is taken as the independent variable, and the expectation degree is used as the dependent variable to start the regression analysis using $m = 0.260x$ ($P < 0.05$). The third step: the school happiness and the expectation degree are determined as independent variables, and the alienation is used as the dependent variable for the



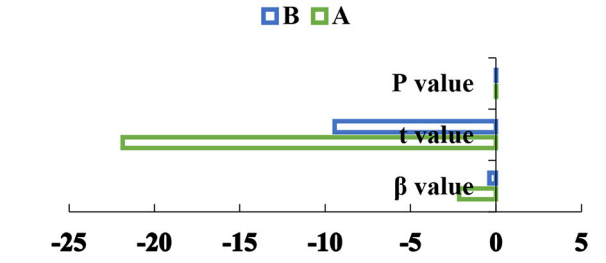
	β value	t value	P value
number	-0.723	-17.24	0.002

FIGURE 7 | The first step of the mediation effect analysis.



	β value	t value	P value
number	0.26	17.87	0.002

FIGURE 8 | The second step of the mediation effect analysis.



	β value	t value	P value
B	-0.374	-9.44	0.002
A	-2.143	-21.85	0.002

FIGURE 9 | The third step of the mediation effect analysis (in the above figures, A indicates that the expectation is the independent variable, and B indicates that the school happiness is the independent variable).

regression analysis using y=-0.374x-2.143m ($P < 0.05$). The specific results are illustrated in **Figures 7–9**.

Figure 7 denotes that the school happiness is selected as the independent variable, and the alienation is selected as the dependent variable for regression analysis, and the *P*-value is 0.002. **Figure 8** refers that the school happiness is selected as the independent variable, and expectation is the dependent variable for regression analysis, and the *P*-value is 0.002. **Figure 9** indicates that the school happiness and the expectation of healthy life in the future are selected as independent variables, and alienation is used as the dependent variable for regression analysis, and the *P*-value is 0.002. It shows a positive correlation when $P < 0.05$.

DISCUSSION

Some experts have pointed out that COVID-19 is an infectious disease of the respiratory tract caused by coronavirus type 2 infection, which can induce severe acute respiratory syndrome. In the world, prevention and control is very difficult (31). Therefore, effective prevention and control publicity and education are needed. The results of this work show that "web browsing" and "social tools (QQ, WeChat, etc.)" are the two most important ways for returning college students to acquire knowledge about the epidemic after the COVID-19 epidemic; in addition, 73.14% of returning college students are very concerned about "the information about the epidemic reported in the media news." The study pointed out that the most trusted source of news is the relevant information released by the official media, and the information stated by the authority has a significant inhibitory effect on the panic of the public (32). There is a positive relationship between the dissemination of information and the risk perception of COVID-19 in the official media (33). It is suggested that schools can use the "school official website, WeChat public account, and QQ group" as well as other platforms to publish publicity and education on the prevention and control of coronavirus disease.

Some documents indicate that alienation is the alienation of the relationship between the individual and the people around him and the society, the natural environment and himself. What's more, it may be controlled and dominated by the object, causing the individual to appear negative emotions such as social isolation, depression, meaninglessness of life, and self-enclosure (34). The average score of returning college students in alienation in physical education class is relatively high. This may be related to everyone's panic about the epidemic, and their lack of knowledge about the epidemic has increased the psychological pressure of returning college students and felt that life is oppressive. Such results are consistent with the findings of Masdrakis et al. (35).

In psychological research, happiness refers to an emotional state that occurs when an individual realizes that his needs are met and his ideals are realized (36). It is a complex and diverse emotional state based on the interaction between psychological factors such as needs (including motivation, desire, and interest), cognition, emotion, and external factors. Happiness indicates the assessor's assessment of the quality of life according to the set standards. The implementation of the overall assessment is a comprehensive indicator to measure the quality of life of an individual (37). In this work, the school happiness of returning college students is relatively high, perhaps because college students have good psychological quality and strong anti-stress ability, can quickly adapt to the environment, and effectively

adjust their negative emotions, so that anxiety can be relieved. The average score of expectations for a healthy life in the future is also relatively high, perhaps because college students live in a relatively simple environment, and their classmates spend a long time with each other and are familiar with each other. In addition, some studies have pointed out that college students have received good higher education, have their own ideas about the future direction, and acquire knowledge relatively quickly, so they have high expectations for future life (38). The disadvantage is that the online questionnaire survey method is used, and the number of respondents is limited and there are regional differences. In future research, a systematic random sampling survey will be conducted according to different regions and different school categories to make the survey results more representative.

CONCLUSIONS

After the COVID-19 pandemic, the cognition on the disease, alienation in physical education class, school happiness, and expectations for a healthy life in the future of Chinese returning college students are investigated and analyzed. It is found that the cognition of college students on COVID-19 is relatively poor. The overall levels of alienation in physical education classes, school happiness, and expectations for a healthy life in the future are relatively high. It can be concluded that the epidemic has a great impact on the psychology of returning college students, and it is necessary to strengthen the mental health education of college students on COVID-19. The disadvantage is that the online questionnaire survey method is used, and the number of respondents is limited and there are regional differences. In future research, a systematic random sampling survey will be conducted according to different regions and different school categories to make the survey results more representative.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Yangzhou University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YX and TL: conceptualization. YT: methodology. SO: software and visualization. YT and TL: validation. YX: formal analysis, investigation, data curation, and writing—original draft preparation. KP and TL: writing—review and editing. KP and SO: supervision. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. Hu Y, Sun J, Dai Z, Deng H, Li X, Huang Q, et al. Prevalence and severity of corona virus disease 2019 (COVID-19): a systematic review and meta-analysis. *J Clin Virol*. (2020) 127:104371. doi: 10.1016/j.jcv.2020.104371
2. Luo H, Tang QL, Shang YX, Liang SB, Yang M, Robinson N, et al. Can Chinese medicine be used for prevention of corona virus disease 2019 (COVID-19)? A review of historical classics, research evidence and current prevention programs. *Chin J Integr Med*. (2020) 26:243–50. doi: 10.1007/s11655-020-3192-6
3. Liu M, Gao Y, Yuan Y, Yang K, Shi S, Zhang J, et al. Efficacy and safety of integrated traditional chinese and western medicine for corona virus disease 2019 (COVID-19): a systematic review and meta-analysis. *Pharmacol Res*. (2020) 158:104896. doi: 10.1016/j.phrs.2020.104896
4. Najar Nobari N, Seirafianpour F, Dodangeh M, Sadeghzadeh-Bazargan A, Behrangi E, Mozafarpoor S, et al. A systematic review of the histopathologic survey on skin biopsies in patients with Corona Virus Disease 2019 (COVID-19) who developed virus or drug-related mucocutaneous manifestations. *Exp Dermatol*. (2021) 30:1233–53. doi: 10.1111/exd.14384
5. Li W, Yang Y, Liu ZH, Zhao YJ, Zhang Q, Zhang L, et al. Progression of mental health services during the COVID-19 outbreak in China. *Int J Biol Sci*. (2020) 16:1732–8. doi: 10.7150/ijbs.45120
6. Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Spijker R, Taylor-Phillips S, et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev*. (2020) 6:CD013652. doi: 10.1002/14651858.CD013652
7. Gibson PG, Qin L, Puah SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. *Med J Aust*. (2020) 213:54–6.e1. doi: 10.5694/mja2.50674
8. Fan E, Beitler JR, Brochard L, Calfee CS, Ferguson ND, Slutsky AS, et al. COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? *Lancet Respir Med*. (2020) 8:816–21. doi: 10.1016/S2213-2600(20)30304-0
9. Mekonnen ZK, Ashraf DC, Jankowski T, Grob SR, Vagefi MR, Kersten RC, et al. Acute invasive rhino-orbital mucormycosis in a patient with COVID-19-associated acute respiratory distress syndrome. *Ophthalmic Plast Reconstr Surg*. (2021) 37:e40–80. doi: 10.1097/IOP.0000000000001889
10. Shelhamer MC, Wesson PD, Solari IL, Jensen DL, Steele WA, Dimitrov VG, et al. Prone positioning in moderate to severe acute respiratory distress syndrome due to COVID-19: a cohort study and analysis of physiology. *J Intensive Care Med*. (2021) 36:241–52. doi: 10.1177/0885066620980399
11. Tsakiri M, Naziris N, Demetzos C. Innovative vaccine platforms against infectious diseases: under the scope of the COVID-19 pandemic. *Int J Pharm*. (2021) 610:121212. doi: 10.1016/j.ijpharm.2021.121212
12. Shen J, Duan H, Zhang B, Wang J, Ji JS, Wang J, et al. Prevention and control of COVID-19 in public transportation: experience from China. *Environ Pollut*. (2020) 266:115291. doi: 10.1016/j.envpol.2020.115291
13. Qaseem A, Etxeandia-Ikobaltzeta I, Yost J, Miller MC, Abraham GM, Obley AJ, et al. Use of N95, surgical, and cloth masks to prevent COVID-19 in health care and community settings: living practice points from the American College of Physicians (Version 1). *Ann Intern Med*. (2020) 173:642–9. doi: 10.7326/M20-3234
14. Cheng SY, Wang CJ, Shen AC, Chang SC. How to safely reopen colleges and universities during COVID-19: experiences from Taiwan. *Ann Intern Med*. (2020) 173:638–41. doi: 10.7326/M20-2927
15. Qiao S, Tam CC, Li X. Risk exposures, risk perceptions, negative attitudes toward general vaccination, and COVID-19 vaccine acceptance among college students in south Carolina. *Am J Health Promot*. (2022) 36:175–9. doi: 10.1177/08901171211028407
16. Alemany-Arrebola I, Rojas-Ruiz G, Granda-Vera J, Mingorance-Estrada ÁC. Influence of COVID-19 on the perception of academic self-efficacy,

state anxiety, and trait anxiety in college students. *Front Psychol*. (2020) 11:570017. doi: 10.3389/fpsyg.2020.570017
17. Santoro E. Information technology e digital health a supporto della salute ai tempi della CoViD-19 [Information technology and digital health to support health in the time of CoViD-19.]. *Recenti Prog Med.* (2020) 111:393–7. doi: 10.1701/3407.33919
18. Pelizza L, Landi G, Pellegrini C, Quattrone E, Azzali S, Pellegrini P, et al. Negative symptom configuration in first episode Schizophrenia: findings from the "Parma Early Psychosis" program. *Nord J Psychiatry*. (2020) 74:251–8. doi: 10.1080/08039488.2019.1695286
19. Moghaddam FM, Hendricks MJ. The psychology of revolution. *Curr Opin Psychol*. (2020) 35:7–11. doi: 10.1016/j.copsyc.2020.02.004
20. Ramsey R, Ward R. Challenges and opportunities for top-down modulation research in cognitive psychology. *Acta Psychol.* (2020) 209:103118. doi: 10.1016/j.actpsy.2020.103118
21. Lo Giudice A, Quinzi V, Ronsivalle V, Martina S, Bennici O, Isola G. Description of a digital work-flow for CBCT-guided construction of micro-implant supported maxillary skeletal expander. *Materials.* (2020) 13:1815. doi: 10.3390/ma13081815
22. Albritton BH, Tonidandel S. How can big data science transform the psychological sciences? *Span J Psychol*. (2020) 23:e44. doi: 10.1017/SJP.2020.45
23. Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, et al. Brief introduction of medical database and data mining technology in big data era. *J Evid Based Med*. (2020) 13:57–69. doi: 10.1111/jebm.12373
24. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M,. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol*. (2020) 21:12. doi: 10.1186/s13059-019-1850-9
25. Sun X, Qin X, Zhang M, Yang A, Ren X, Dai Q. Prediction of parental alienation on depression in left-behind children: a 12-month follow-up investigation. *Epidemiol Psychiatr Sci*. (2021) 30:e44. doi: 10.1017/S2045796021000329
26. Wu S, Zheng X. The effect of family adaptation and cohesion on the happiness of married women: a multiple mediation effect. *J Gen Psychol*. (2020) 147:90–107. doi: 10.1080/00221309.2019.1635075
27. Lee SM, Yoo JI, Youn HS. Changes in alienation in physical education classes, school happiness, and expectations of a future healthy life after the COVID-19 pandemic in Korean adolescents. *Int J Environ Res Public Health*. (2021) 18:10981. doi: 10.3390/ijerph182010981
28. Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes Metab Syndr*. (2020) 14:1467–74. doi: 10.1016/j.dsx.2020.07.045
29. Ye Y, Shi J, Zhu D, Su L, Huang J, Huang Y. Management of medical and health big data based on integrated learning-based health care system: a review and comparative analysis. *Comput Methods Programs Biomed*. (2021) 209:106293. doi: 10.1016/j.cmpb.2021.106293
30. Chen Y, Han D, Wang J, Yang Y, Hu W. Stacking of micro-aperture arrays: a new strategy to construct Söller collimator for x rays. *Rev Sci Instrum*. (2020) 91:073109. doi: 10.1063/5.0010978
31. Safiabadi Tali SH, LeBlanc JJ, Sadiq Z, Oyewunmi OD, Camargo C, et al. Tools and techniques for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19 detection. *Clin Microbiol Rev*. (2021) 34:e00228–20. doi: 10.1128/CMR.00228-20
32. Zhang Y, Akhtar N, Farooq Q, Yuan Y, Khan IU. Comparative study of Chinese and American media reports on the COVID-19 and expressions of social responsibility: a critical discourse analysis. *J Psycholinguist Res*. (2021) 9:1–18. doi: 10.1007/s10936-021-09809-9
33. Chao M, Chen X, Liu T, Yang H, Hall BJ. Psychological distress and state boredom during the COVID-19 outbreak in China: the role of meaning in life and media use. *Eur J Psychotraumatol*. (2020) 11:1769379. doi: 10.1080/20008198.2020.1769379
34. McMillan DW, Lorion RP. Sense of community, pathway to, or bridge from alienation? *J Community Psychol*. (2020) 48:1706–14. doi: 10.1002/jcop.22406
35. Masdrakis VG, Legaki EM, Papageorgiou C, Markianos M. Psychoticism in patients with panic disorder with or without comorbid agoraphobia. *Int J Psychiatry Clin Pract*. (2017) 21:181–7. doi: 10.1080/13651501.2017.1305111
36. Alexander R, Aragón OR, Bookwala J, Cherbuin N, Gatt JM, Kahrilas IJ, et al. The neuroscience of positive emotions and affect: implications for cultivating happiness and wellbeing. *Neurosci Biobehav Rev*. (2021) 121:220–49. doi: 10.1016/j.neubiorev.2020.12.002
37. Choi I, Yu J, Lee J, Choi E. Essentializing happiness reduces one's motivation to be happier. *J Pers*. (2021) 89:437–50. doi: 10.1111/jopy.12589
38. Al-Hinaai H, Al-Busaidi I, Al Farsi B, Al Saidi Y. The prevalence of substance misuse and its effects among omani college students: a cross-sectional study. *Oman Med J*. (2021) 36:e224. doi: 10.5001/omj.2021.04

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xue, Pyong, Oh, Tao and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fpubh.2022.917330

Application Research of Public Art in the Design of Old People's Rehabilitation Space Based on Virtual Information Interaction Platform From the Perspective of Developing Economies

*Shaoqing Zhang**

College of Landscape Architecture and Art, Henan Agricultural University, Zhengzhou, China

OPEN ACCESS

Edited by:
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Yalin WU,
Lushan Botanical Garden (CAS), China
Zhuang Wu,
Capital University of Economics and Business, China

***Correspondence:**
Shaoqing Zhang
shaoqing7615@sina.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *11 April 2022*
Accepted: *03 May 2022*
Published: *27 May 2022*

Citation:
Zhang S (2022) Application Research of Public Art in the Design of Old People's Rehabilitation Space Based on Virtual Information Interaction Platform From the Perspective of Developing Economies. Front. Public Health 10:917330. doi: 10.3389/fpubh.2022.917330

To solve the problem of the design of the old people's recuperation space, the virtual information interaction platform is used to study the public art application in the design of the old people's recuperation space. Firstly, the principles of interactive design are expounded, and secondly, the existing institutions for the old people are investigated. Under the premise of optimizing the functions of the facilities, the concepts of humanistic care, emotional care and humanization in public art are integrated into the design of the old people's rehabilitation space, to solve the long-term negative impression of the old people's repression and indifference to the old people's care institutions. The construction of the scene allows the old people to experience some operations with the help of the virtual information interaction platform. In the modern elderly rehabilitation space, the attention and application of public art design will inevitably bring spiritual and material help to the old people in their later years, and create a happy, peaceful, and comfortable elderly life for them. The survey results manifest that 65.3% of urban old people and 71.8% of rural old people feel that they cannot keep up with the pace of development. Through the analysis and discussion of the physiological and psychological characteristics of the old people, the whole survey denotes that the physiological functions of the old people are declining, which seriously affects their normal life. Therefore, the design of the rehabilitation space for the old people should not only meet the basic needs of life, but also analyze the space design from the perspective of humanization and emotion. An ecological, natural, and human settlement environment has been established. The recuperation space is designed for the needs of different old people, which helps the old people to eliminate loneliness, enhance their value of the old people, and make life full of joy and meaning for the old people.

Keywords: old-age, recuperation space, virtual information interaction platform, public art, design

INTRODUCTION

The aging of the population is gradually becoming a worldwide topic and an important issue for many countries at this stage. Since the end of the last century, China has entered an aging society. China is a country with a large number of old people population in the world, and also one of the countries with the fastest growing degree of aging (1, 2). How to deal with the needs of many old people and provide more comfortable living conditions for older people is one of China's current national conditions. From the change of social and family roles, the old people are prone to psychological imbalance, which leads to loneliness. In addition to caring for their physical health, the focus of the old people's emotional space is equally critical, so that their later life can be full of happiness and care. In China, with the rapid growth of the elderly population and a large base, little attention has been paid to the old people, and less care has been given to the old people (3). After the old people withdrew from the social role, most of the time they lived at home or in the old-age institutions, and they were more dependent on their living space in their later years. They need to go to public life and communicate with others. Therefore, the living space of the old people should create a living space suitable for health, recreation, emotional communication, social play, and pleasing life (4).

There are many studies on the old people. For example, Lee et al. (5) developed eight personas representing four different user groups in the context of home appliance use: visually impaired, hearing impaired, spinal cord impaired, and the old people. Each has a persona as a tool to understand the target user, a persona card representing their task obstacles, frustrations, needs and references, and a cartoonish persona. In this study, two common accessibility issues and two role-specific issues are addressed in each user group (5). These questions are presented in the language roles of stakeholders to help them understand users. Tsuchiya et al. (6) conducted a study on the evaluation of smart home control systems for the old people with a sample of 10 users in inland cities in Brazil. The control system consisted of a prototype using a web-based mobile application, which was developed taking into account requirements obtained from previous studies and recommendations from the literature The study demonstrated clearly the importance of providing information directly by identifying the user's home location in the application, not hiding information under scrolling, appropriately using images and videos in the system to avoid confusion, limiting the number of open windows to maintain context, avoiding unclear interactive elements to facilitate, and showing the visuals of the room and appliances on the screen (6). The findings could help improve the old people's interactions with smart home systems, especially in rural areas of developing countries. And public art is no longer an unfamiliar word in contemporary China. After the reform and opening up in the 20th century, it became a topic of great concern to Chinese artists in the 1990s (7). The development of contemporary Chinese public art is gradually developed and improved along with the adjustment of national policies, the marketization and internationalization of the economy, the socialization of rapid urban development, and the continuous improvement of citizens' rights. Since the beginning of the 21st century, public art has shifted from intense topic discussion to practical research in China, and is entering the urgent needs and rapid development of contemporary society (8, 9). With the rapid development of China's economy and the urgent need for public art culture, China's contemporary public art and its theories are constantly advancing and exploring, and are influenced by relevant external experience and theoretical research. Combined with the current situation of Chinese society and the needs and problems of public art culture development, corresponding public art practices and theories have gradually been produced (10). The main feature of its theoretical research is to gradually shift from the study of the external form and style of public art and the application of urban public space to the study of public art and sociology. The study of the simple aesthetic value orientation of public art has been transformed into the study of the relationship between public art and urban subject culture and citizens' daily public life. Contemporary Chinese public art no longer exists in only one form of art. Nowadays, the publicity and public participation of public art are emphasized. Various services are provided to the public and diverse forms of social public space, which is the overall social and cultural mission. At present, it pays more attention to various vulnerable groups in the society and realizes social fairness and democratic rights in the form of art (11). Therefore, based on Chinese contemporary public art, the recuperation space for the old people is studied.

With the improvement of people's living standards, its actual functions should be fully considered in the design, to meet the needs of the old people for material life as much as possible. According to the survey, it is found that most of the old people lack communication, have a sense of distance, and lack communication bonds with each other. Through the reasonable planning of public space, it is the main motivation of this research to make the people in the space connect with each other and the environment, and make more contacts. The old people have many needs on the spiritual level, which can be roughly divided into spiritual needs such as emotional, social, entertainment and self-realization. Today, the existing old-age care places in the society can no longer fully meet the spiritual needs of the old people, and the design also needs to be improved. The main innovation is to solve the spiritual needs of the old people. It is designed from three aspects: social, participatory, and service-oriented. Interactivity is applied to the layout of the public space of the old people apartment to solve the needs of the old people. This research explores how to carry out the interactive design in the design of rehabilitation space, enhance the ability of the old people's care services, improve the service level for the old people, and enrich the theory of social elderly care, which has good research value.

METHODOLOGY

The Outdoor Environment Design of the Rehabilitation Space for the Old People

The outdoor environmental space refers to the spatial extent of the specific area except the interior of the building, which is

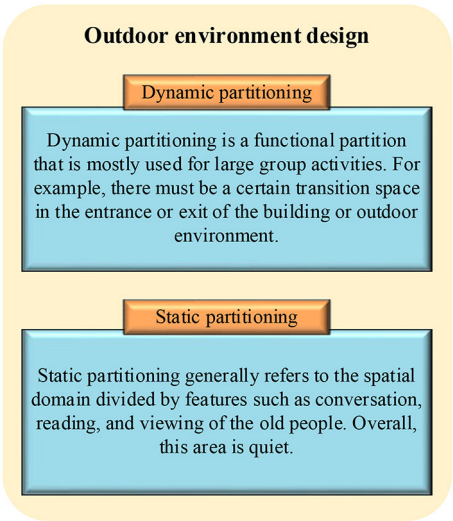


FIGURE 1 | The outdoor environment design of the elderly rehabilitation space.

different from the open space of the city (12). The old people in nursing homes have different cultural levels and different physical qualities. Therefore, in order to meet these different needs of the old people, the way is also ever-changing (13). The overall differences in the activities of different types of elderly groups and the different characteristics of their activities have become the basis for the outdoor environment design in the elderly rehabilitation space.

Necessity, spontaneity, and sociality are several environmental categories in the aged rehabilitation space. The main activities in the entire elderly rehabilitation space are mainly based on essential and spontaneous activities. All activities of the old people are particularly dependent on the quality of the environment and space. When the quality of external activity space is relatively bad, old people mainly focuses on activities that are essential to life, while other interactions are basically not going on. However, when the quality of the activity space environment is good, they will spontaneously generate diverse activities, and social activities will increase.

When setting up the activity space, special consideration is given to the physiological characteristics of the old people. Ventilation and lighting must be very good. However, the outdoor activities of the old people cannot be located at the vents, and the outdoor rest points for the old people should be continuous. The old people have limited physical strength and the outdoor seating should be installed along the road.

The outdoor environment design of the rehabilitation space for the old people is divided into two parts, as shown in **Figure 1**. Outdoor environment design is usually divided into dynamic partitioning and static partitioning. Static partitioning generally refers to the spatial domain divided by features such as conversation, reading, and viewing of the old people. Overall, this area is quiet. Dynamic partitioning is a functional partition that is mostly used for large group activities. For example, there must be a certain transition space in the entrance or exit of the building or outdoor environment. Seats and ornamentals for leisure must be provided. The composition and diversity of space is emphasized when designing outdoor spaces. Light, shadow, and visual effects should be appropriate for older people's life requirements, which avoids the overly shocking picture and space composition.

Functional Composition of the Elderly Rehabilitation Space

The functional composition of the elderly rehabilitation space is shown in **Figure 2**.

Accommodation Function

The accommodation function is the most important requirement of the old people in the rehabilitation space. Old people generally spend most of their time in the living space. Therefore, the accommodation environment for the old people must be taken very seriously. According to the physiological characteristics of the old people and the requirements of life use, the necessary functional facilities, such as the bathroom, storage space, balcony area, emergency equipment, etc., should be considered in the design. At the same time, privacy requirements should also be considered.

Leisure Function

In rehabilitation space, the problem of rest must be involved, whether indoor or outdoor. In the discussion of rehabilitation space, the external environment is the main design object. Therefore, in the environment of external communication activities, reasonable rest places must be set up. Old people often need to sit and rest during outdoor activities for physical reasons. Therefore, in the external landscape area, road walking area, leisure activity area, building periphery and other important occasions, leisure seats should be set up, and the number is higher than the design ratio in the city square. At the same time, the ergonomic requirements of the old people should be met.

Health Care Function

Old people's physical condition is much worse than before, especially the elderly people. Sudden infirmities of age often occur and emergency medical care is needed. Therefore, in the rehabilitation space, the design of the medical function is very important. This guarantees the safety of the old people in the event of an emergency. In the design of the location of the medical space, the road must be smooth and convenient, so the ambulance can arrive as fast as possible, and the rescue time is shortened (14).

Barrier-Free Design Requirements

Barrier-free design is an important factor that must be considered and embodied in rehabilitation space for the elderly. Barrier-free design helps old people perform normal activities. This assists to solve the problems of old people, such as the decline of physiological function and inconvenience of movement. Barrier-free design should be strictly observed in all parts of rehabilitation space for the elderly and in the placement of functional facilities (15).

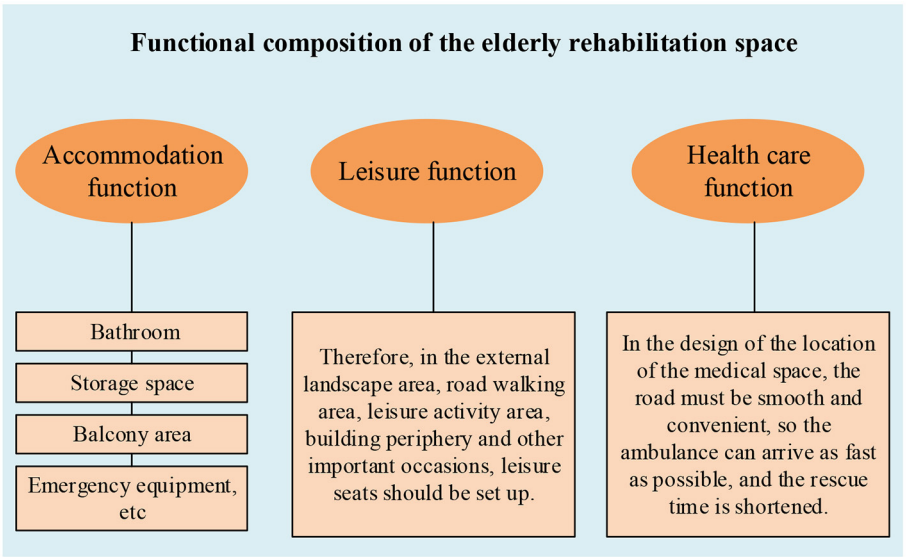


FIGURE 2 | The functional composition of the elderly rehabilitation space.

The Influence of Old People's Physical Decline

Design of the ramp: In the rehabilitation space of the elderly, it is inevitable that there will be road ramps, including stairs with a ground height difference environment. Therefore, the design of the ramp should be based on gentle slopes as much as possible. It is convenient for wheelchairs and other equipment to walk, and must be equipped with safety rails next to it. In addition, there should be reasonable ramps side by side in places where there are steps.

Use of Non-slip Materials

The balance ability of the old people and the flexibility of the limbs are bad, so the floor tiles must be strictly controlled. Floor coverings with high surface smoothness are strictly prohibited, which makes it easy for the elderly to slip. Especially in corridors, stairs, sanitary napkins, etc., these are places where old people are prone to accidents. In outdoor activities, paving should be based on flat and non-slip materials.

The Influence of Hearing and Vision of Old People

After the decline of physiological function, the first manifestation is the change of hearing and vision. Therefore, in the maintenance of the modern elderly, in view of the hearing loss of the old people, there are different amplification devices in different space areas to ensure the normal activities and exchanges of the elderly. For the decline of vision, the following aspects are considered.

Eye-catching signs are set on the road and orientation search, and distinguished by color and shape. Good lighting systems are equipped to assist in lighting. Through voice control and other means, these lighting devices are connected, especially at night. Intelligent devices, some acousto-optic signal devices, such as telephones, doorbells, elevator stop signals, call service facilities, etc. should be strictly observed in the design of the indicator lights. This is brighter than the usual general equipment and is clearly legible for identification. The intelligent guiding device is configured to be carried by the elderly. Whether indoors or outdoors, the location of the elderly can be searched at any time to guide the direction, which helps the visually obsessed old people to live comfortably.

Management Discussion

A sound management system for the elderly rehabilitation space was established. The professional knowledge and skills training of nursing staff was strengthened and the quality of personnel services was improved (16). Moreover, the state should have a strong and sound legal system supervision. The healthy life of the old people is advocated. Self-care ability is improved through the efforts and exercise of the old people themselves. Helping the old people to rebuild their confidence is also one of the important purposes of the construction of the elderly rehabilitation institutions.

Physiological Characteristics of the Elderly and Its Impact on Public Art Design in the Rehabilitation Space of the Elderly

The physiological function of the human body will continue to decline, which is the objective law of human life development. However, the decline of physiological functions will directly affect the lives of the old people and their ability to adapt to the natural environment and living environment. In addition, there will be a significant decline in communication skills. Therefore, in the design of the rehabilitation space for the elderly, the design of the living space and the outdoor environmental landscape should be based on the actual situation of the elderly.

The process of the old people's decline is divided according to the "old people's residential building design guidelines", which

are mainly in four stages: The healthy functioning period of the body function is 60–64 years old. The independent period of self-care is 65–74 years old. The inconvenience of action is 75–84 years old. 85 years of age or older is a serious debilitation and is listed as a key nursing period.

The normal growth and development of the human body will change over time. As the age increases, the overall signs of physiologic weakness are more pronounced. The growing age has gradually degraded the functions of the various parts of the old people's body. Especially after 70 years of age, the signs of this decline are more obvious, which will seriously affect the normal physiological function of the old people. As physiology decreases, the design tools and methods used for the needs of the old people will change. More requirements are also placed on the design of public art in the aged rehabilitation space. According to the special requirements of the old people's physiological functions, public environmental facilities or public works of art that meet the overall requirements are designed.

The degree of self-care of the old people directly determines the degree of privacy of the entire living space and the level of living services and care that the maintenance center needs to provide to the old people. Under normal circumstances, the continuous decline of the old people's self-care ability will directly lead to the weakening of the privacy of the entire old people's living environment. At the same time, the maintenance center will be required to strengthen the care and companionship of the old people who cannot take care of themselves. The entire care phase of an old person consists of the following steps: life self-help period; partial care period; comprehensive care period; termination period. From self-help to termination, the life of the old people is an unchangeable trend. The level of care varies according to the difference in care required at each stage of human normal aging.

In combination with the differences in the physiological characteristics of the older people mentioned above, there is a big difference between adults and old people. The physiology of older elderly people is more fragile. Therefore, more care is needed for older people, considering the particularities of life of older people. Thus, there are several factors that must be considered when designing the old people rehabilitation space.

Auditory System Design

A very prominent feature of the old people group is the sharp deterioration of hearing due to the decline of physiological functions. In the investigation of old people rehabilitation institutions, there have been many cases in which the old people have difficulty communicating due to severe hearing loss. Old people are eager to have sensitive ears to get useful information in life. A sharp decline in hearing can significantly reduce the enthusiasm of older people to participate in activities. Older people will have obvious conversational conflicts in their hearts, and they need tools to help their ears get information.

However, when the old people are alone, they like quiet places and spaces very much. They don't want their quiet state to be disturbed by loud noises. In a survey of nursing homes, it was found that old people were disturbed by noisy external voices because it lived close to the main road. Therefore, for the phenomenon of hearing loss of the old people, the design of the elderly rehabilitation space also brings a lot of challenges.

Vision System Design

A decrease in vision is an important manifestation of the decline in the physiological function of the old people. In previous interviews with several elderly institutions, it was learned that many problems, such as dazzling eyes and difficulty in light perception, will arise because of the impaired vision. The old people went in the wrong direction and even stumbled because they could not see the marker. Therefore, in the design of the elderly rehabilitation space, the visual situation of the old people must be carefully understood. The subject buildings, road-pointing signs, and indexing of important places involve visual observations of the old people, and any environment must be treated with rigorous care. For example, a very prominent reminder sign needs to be placed on the floor building, the stairway, the corner, the slope, and so on.

Tactile System Design

Old people's tactile system is also declining with age. They rely on their physical touch to feel the environment and place very high demands on the environment. For cold and hot, the old people are significantly less able to accept the environment than young people. For example, the old people like to sit around the fire source in winter, or need heating facilities. In summer, they need a cool and ventilated environment. Therefore, there are many requirements for the design of the old people rehabilitation space, which is different from the treatment of young people. At present, the design of old-age maintenance institutions is seriously inadequate. In the future, the elderly rehabilitation institutions have higher requirements.

Barrier Free Design

The concept of barrier-free design must be strictly enforced throughout the design of the elderly rehabilitation space. The old people will continue to decline in their physiological functions, their physical activities will be inconvenient, and the ability to adapt to the environment will show significant decline. Steps, stairs, corridors, handrails, toilets and other very common public facilities, young people usually do not pay much attention to. However, as the physical and mental functions continue to decline, the old people will have inexplicable fears and potential threats to these common public facilities. In the course of practical research, it was found that most of the old people needed crutches and other supporting tools in their activities.

Activity Space Design

As a result of the comprehensive analysis of the physiological characteristics of the old people, as their age increases, the functional organs will also weaken. The overall health status will be significantly lower than in the past. Therefore, the old people will have a very personal health status in their lives, and they are also very concerned about the disease.

From the analysis of physiological characteristics, the old people need a public green space with good environmental quality to breathe fresh air and exercise. In addition, the old people have come out of the life stage of youth and live in

society. However, they have withdrawn from many social things such as social interaction. Second, the dominant role of the old people in society is also changing anyway. From the collective work of the unit to the individual work in the family, the entire social environment and social group relations are also changing. Children lived in other places all the year round and stayed away from the elderly. The lively atmosphere of the family had changed. Therefore, the new living environment, life progress and lifestyle will bring a sense of discomfort. These inevitably lead to the feeling of anxiety and loneliness in the old people. Compared with some families with superior economic conditions, the high-cultural and healthy old people will have great enthusiasm and interest in social, group entertainment and social interaction. This type of old people is hoping that society can provide them with a place that is relatively safe and can satisfy their social interaction activities. Combined with the physical characteristics of the old people, the design of activity venues must include natural health factors, which reflects a kind of care and consideration for old people. The event space for the elderly has a reasonable radius to meet the needs of various entertainment activities. It creates a healthy, comfortable, self-cultivation and physical exercise ecological environment for the old people.

RESULTS AND DISCUSSION

Psychological Characteristics of the Elderly and Their Impact on Public Art Design in the Elderly Rehabilitation Space

Because of changes in the living environment, the social roles and family roles of the old people will also change, which will have a very significant impact on the psychology of the old people. At the same time, the mental state of the old people has changed a lot because of the different nature of the work, the degree of education, the differences in local culture, and the differences in social experience. In the survey, it was found that loneliness, impatience, survival, resentment, attachment, symbiosis, etc. are the common psychological feelings of the old people.

Loneliness

Many old people's spouses may die, the circle of life and activities is smaller, and the number of people who can communicate with themselves is less. In addition, the phenomenon of "empty nest" appears. Old people feel lonely when their children are not around. Therefore, most old people are reluctant to communicate, like to be alone, and not willing to participate in group activities. Lonely feelings can have a very negative impact on the lives of the old people.

Lost Feeling

Lost feeling is also a situation that the old people often encounter. Old people feel left out and abandoned. This psychological pressure caused some of the old people to change from self-care to dependents. The sense of loss seriously affects the normal living conditions of the old people, which can cause mental tension and frustration in the old people.

Inferiority

In the elderly rehabilitation space, the old people will have a feeling of inferiority because of the difference between themselves and others. Especially when the ability is gradually declining and cannot reach a good state, the old people will inevitably have a tendency to feel inferior, and always think in the subconscious that they are no longer able to do so. The inferiority of old people is also a common occurrence. Because of retirement, social status and economic income are gone. At the same time, the physical condition is not as good as before, the elderly need to be taken care of by others. Therefore, they will have a feeling of inferiority.

Isolation

Isolation is the most common phenomenon in the old people group, and it is also a very realistic and serious situation. The very important reason for the emergence of these phenomena is that there is no comfort in the life of the old people. Spiritual comfort cannot be satisfied and cannot make an understanding of the spiritual world. Therefore, ideologically, the elderly will create pressure on themselves, and eager for the care and attention of the people around them, especially their relatives and friends. In view of this, in the design of the rehabilitation space for the elderly, it is necessary to solve the problem of the public art design in the space. At the same time, the accurate grasp of the psychological characteristics of the old people is of great help to design and construction of the old facilities and the public art of the elderly rehabilitation space.

Through investigation on some aspects, such as the urban elderly and the rural elderly cannot keep up with the development of society, the old people are the burden of the society and the family, the society is paying more and more attention to the importance of the old people, the old people like to chat with others and make friends, and the old people often feel lonely. The results are shown in **Table 1**. According to a sample survey, 65.3% of urban old people and 71.8% of rural old people feel that they cannot keep up with the pace of development. 52.5% of urban old people and 61.2% of rural old people consider themselves to be social burdens. It can be seen that the rural old people are far less than the urban old people in their recognition of their own values. Therefore, for the old people in township nursing homes, they will feel that they can't keep up with the pace of development of the times, and they are the burden of family and society. On the emotional issues of the old people, the urban old people who like to chat with others account for 70.9%, and the rural old people account for 76.2%. In terms of making friends, the urban old people who choose to make friends are 63.7%, and the rural old people account for 59.1%. It is also worth noting that the rural old people often feel lonely more often than the urban old people. In short, the old people, especially the rural old people, consider themselves to be the burden of society and cannot keep up with the development of society. Psychological problems are more serious. It is conceivable that this situation will be more serious for the old people of the nursing home, especially in terms of emotional comfort.

Various psychological characteristics were studied. Old people will choose different activities due to different psychological characteristics. Changes in the mental state of the old people

TABLE 1 | Composition of various statements of the elderly in the surveyed cities and rural areas.

	City			Rural		
	Yes	Unclear	No	Yes	Unclear	No
Can't keep up with the development of society	65.3%	13.3%	21.4%	71.8%	14.3%	13.9%
Old people are the burden of society	52.5%	16.4%	31.1%	61.2%	19.8%	19.0%
Old people are the burden of the family	46.9%	15.1%	38.0%	67.3%	15.6%	17.1%
The society is increasingly concerned about the importance of the elderly	82.1%	12.3%	5.6%	86.2%	10.5%	3.3%
Older people like to chat with others	70.9%	10.3%	18.8%	76.2%	9.1%	14.7%
Older people like to make friends	63.7%	12.8%	23.5%	59.1%	15.3%	25.6%
Old people often feel lonely	24.9%	11.4%	63.7%	37.6%	16.1%	46.3%

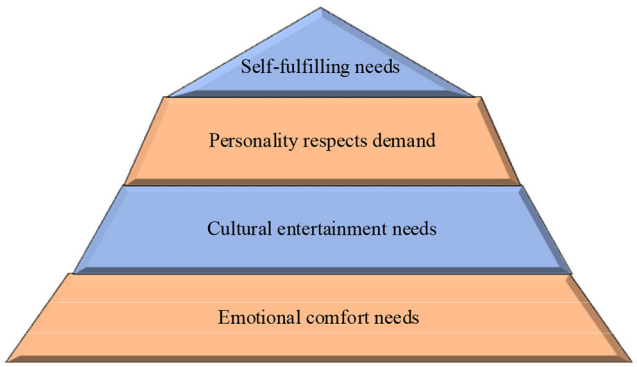


FIGURE 3 | The spiritual needs of the elderly.

will directly affect the design of public art in the old people rehabilitation space. Therefore, the factors influencing this aspect must always be considered in the design. At present, in the elderly rehabilitation space, there are relatively few design studies on the psychological characteristics of the old people. The spiritual needs of the elderly are shown in **Figure 3**.

Safety

The body function of the old people is declining and the movement is inconvenient. The ability to recover from injuries is also weakening. Therefore, security guarantees are important. Environmental design and landscape design in the public space of the elderly must be taken seriously. For example, barrier-free design, emergency call system, emergency system and medical security system are all considerations of old people security requirements.

Sense of Belonging

When entering a new living environment and state, the old people need to be recognized by the people and the environment around them to enhance their self-confidence. In this way, these old people can integrate into the collective, build confidence and be accepted and recognized. A sense of belonging is also a sense of responsibility of the old people to the environment and the people around them. Old people who lack a sense of belonging will have a clear lack of responsibility and vice versa.

TABLE 2 | Analysis of scope of activities.

Scope of activities	Percentage
Home	16%
Community	20%
Square	32%
Elderly activity room	9%
Outdoor	23%

Neighborhood

The neighborhood of the old people is mainly reflected in the social neighborhood and the environmental neighborhood. Social neighborhood means that the old people need to find friends in the new environment who can communicate, exercise and play games with themselves. An environmental neighborhood means that the old people are intimate with the environment of life and are willing to live in this environment. When designing the living space of the old people, the design of the communication environment of the entire event space was improved, and a new neighborhood communication space was created.

Family Sense

In the design of the elderly rehabilitation space, in addition to considering the psychological feelings of the old people, the sense of family also needs to be considered. The family atmosphere is created. The old people in the old-age care institutions have left the familiar life field that has lived for decades, and it takes time to adjust the lever for the new living environment. The sense of family in the whole environment depends on the accumulation of time.

Public Space for the Old People's Convalescence

The public space for elderly rehabilitation is mainly divided by the space range, as shown in **Table 2**. After a survey, it is found that most people aged 50-70 like to live in public spaces such as communities, squares, activity rooms and outdoors. Therefore, the scope of this design is divided into two parts: indoor space and outdoor space. The indoor public space includes a leisure reading area, display area, reading room, fitness room, yoga room and performance hall. In the outdoor public space, there are

rooftop leisure gardens, landscape viewing spaces, and fitness spaces. Indoor public space and outdoor public space together constitute the public space category of this research.

Discussion of Results

Through the analysis and discussion of the physiological and psychological characteristics of the old people, the entire survey shows that the physiological functions of the old people are declining, which seriously affects their normal life. Psychologically, old people are also suffering from a variety of social and family role changes. A healthy mindset can help the old people live a happy life, but most of the old people have a hard time accepting the current situation, from the main role of society to retirement. The main difficulty of the old people is to face a new living environment and establish a new personal relationship and emotional space. Therefore, the design of the old people rehabilitation space should analyze the spatial design from the perspective of humanization and emotionality while meeting the basic needs of life. An ecological, natural and humane living environment has been established. In the old people rehabilitation space, public art is designed to fully consider their needs. In the rehabilitation space, older people can spend their old age in happiness. The quality of the living environment is good. Furthermore, connecting people through shared experiences and emotions focuses on connecting people to artworks and connecting them to specific communities, which is consistent with the work of Zebracki et al. (17).

CONCLUSION

Today, society is facing an increasingly serious phenomenon of aging, and old people are getting more and more attention. The basic starting point of public art, public environment, and public facilities in the elderly rehabilitation space is to meet the physical and psychological needs of the old people, to facilitate their use of the old people, and to provide a comfortable living environment. At the same time, in order to reduce the psychological pressure of the old people, a healthy, harmonious and happy rehabilitation space was established. However, the current main research task for the construction of old-age care institutions should be based on the psychological and behavioral characteristics of the old people. The society is constantly improving, and the concept of life of the old people in each era is constantly changing and continuing. Therefore, the psychological research on the old people is constantly changing, and the design concept of the elderly rehabilitation space will be constantly changing and updating. Taking the elderly as the research object, their physiological, psychological and behavioral characteristics were studied. Existing geriatric care institutions have been investigated on the spot, which provides relatively intuitive research data for exploring conclusions. The results show that the recuperation space design should be built according to the needs of different old people, which helps the old people to eliminate loneliness. At the same time, to enhance the value of the role, the old people can also make life full of joy and meaning. The shortcoming is that it only analyzes the feasibility of the application of public art experience design for the old people. However, the interactive behavior of the old people's care space is related to each other and affects each other. Therefore, a public space that can give the old people a sense of belonging and comfort should be created as much as possible to meet their psychological and physical needs. Creating a public space environment full of communication and interaction in the rehabilitation space for the old people can not only provide new design ideas and research directions for the public space design of elderly apartments, but also enrich other fields of environmental art design in theory and practice.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki and the protocol was approved by the Ethics Committee of Wonkwang University (Project Identification Code 13843/CNA).

AUTHOR CONTRIBUTIONS

SZ: writing—original draft preparation and review and editing.

FUNDING

This work was supported by the China Postdoctoral Science Foundation.

REFERENCES

1. Horak F, King L, Mancini M. Role of body-worn movement monitor technology for balance and gait rehabilitation. *Phys Ther.* (2015) 95:461–70. doi: 10.2522/ptj.20140253
2. Beaulieu-Boire L, Belzile-Lachapelle S, Blanchette A, Desmarais PO, Lamontagne-Montminy L, Tremblay C, et al. Balance rehabilitation using Xbox Kinect among an elderly population: a pilot study. *J Nov Physiother.* (2015) 5:261. doi: 10.4172/2165-7025.1000261
3. De Coninck L, Bekkering GE, Bouckaert L, Declercq A, Graff MJ, Aertgeerts B. Home-and community-based occupational therapy improves functioning in frail older people: a systematic review. *J Am Geriatr Soc.* (2017) 65:1863–9. doi: 10.1111/jgs.14889
4. Levasseur M, Généreux M, Bruneau JF, Vanasse A, Chabot É, Beaulac C, et al. Importance of proximity to resources, social support, transportation and neighborhood security for mobility and social participation in older adults: results from a scoping study. *BMC Public Health.* (2015) 15:503. doi: 10.1186/s12889-015-1824-0

5. Lee JH, Kim YM, Rhiu I, Yun MH. A persona-based approach for identifying accessibility issues in elderly and disabled users' interaction with home appliances. *Appl Sci.* (2021) 11:368. doi: 10.3390/app110 10368
6. Tsuchiya LD, Braga LF, de Faria Oliveira O, de Bettio RW, Greghi JG, Freire AP. Design and evaluation of a mobile smart home interactive system with elderly users in Brazil. *Pers Ubiquitous Comput.* (2021) 25:281–95. doi: 10.1007/s00779-020-01408-0
7. Barker AL, Talevski J, Bohensky MA, Brand CA, Cameron PA, Morello RT. Feasibility of Pilates exercise to decrease falls risk: a pilot randomized controlled trial in community-dwelling older people. *Clin Rehabil.* (2016) 30:984–96. doi: 10.1177/0269215515606197
8. Brach JS, Francois SJ, VanSwearingen JM, Gilmore S, Perera S, Studenski SA. Translation of a motor learning walking rehabilitation program into a group-based exercise program for community-dwelling older adults. *PM R.* (2016) 8:520–8. doi: 10.1016/j.pmrj.2015.10.004
9. Nordbakke S, Schwanen T. Well-being and mobility: a theoretical framework and literature review focusing on older people. *Mobilities.* (2014) 9:104–29. doi: 10.1080/17450101.2013.784542
10. Dahlkvist E, Hartig T, Nilsson A, Högberg H, Skovdahl K, Engström M. Garden greenery and the health of older people in residential care facilities: a multi-level cross-sectional study. *J Adv Nurs.* (2016) 72:2065–76. doi: 10.1111/jan.12968
11. Foster M, Allen S, Fleming J. Unmet health and rehabilitation needs of people with long-term neurological conditions in Queensland, Australia. *Health Soc Care Community.* (2015) 23:292–303. doi: 10.1111/hsc.12146
12. Magasi S, Wong A, Gray DB, Hammel J, Baum C, Wang CC, et al. Theoretical foundations for the measurement of environmental factors and their impact on participation among people with disabilities. *Arch Phys Med Rehabil.* (2015) 96:569–77. doi: 10.1016/j.apmr.2014.12.002
13. Rantakokko M, Portegijs E, Viljanen A, Iwarsson S, Kauppinen M, Rantanen T. Changes in life-space mobility and quality of life among community-dwelling older people: a 2-year follow-up study. *Qual Life Res.* (2016) 25:1189–97. doi: 10.1007/s11136-015-1137-x
14. Portegijs E, Rantakokko M, Mikkola TM, Viljanen A, Rantanen T. Association between physical performance and sense of autonomy in outdoor activities and life-space mobility in community-dwelling older people. *J Am Geriatr Soc.* (2014) 62:615–21. doi: 10.1111/jgs.12763
15. Polku H, Mikkola TM, Portegijs E, Rantakokko M, Kokko K, Kauppinen M, et al. Life-space mobility and dimensions of depressive symptoms among community-dwelling older adults. *Aging Ment Health.* (2015) 19:781–9. doi: 10.1080/13607863.2014.977768
16. Kammerlind ASC, Fristedt S, Ernsth Bravell M, Fransson EI. Test–retest reliability of the Swedish version of the Life-Space Assessment Questionnaire among community-dwelling older adults. *Clin Rehabil.* (2014) 28:817–23. doi: 10.1177/0269215514522134
17. Zebracki M, Xiao M. The landscape of public art research: a knowledge map analysis. *Prof Geogr.* (2021) 73:481–95. doi: 10.1080/00330124.2021.1878907

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fpubh.2022.916364

A Prognostic Model for Colon Adenocarcinoma Patients Based on Ten Amino Acid Metabolism Related Genes

Yangzi Ren [1†], Shangwen He [2†], Siyang Feng [1,3] and Wei Yang [1,4*]*

[1] Department of Pathology, Guangdong Province Key Laboratory of Molecular Tumor Pathology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China, [2] The First School of Clinical Medicine, Southern Medical University, Guangzhou, China, [3] Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China, [4] Research Department of Medical Sciences, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Sravan Kumar Reddy,
R G M College of Engineering and Technology, India
Rmahammad Shafi,
Mizan Tepi University, Ethiopia

***Correspondence:**
Yangzi Ren
R13288649572@163.com
Wei Yang
yanglabgz@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 09 April 2022
Accepted: 03 May 2022
Published: 27 May 2022

Citation:
Ren Y, He S, Feng S and Yang W (2022) A Prognostic Model for Colon Adenocarcinoma Patients Based on Ten Amino Acid Metabolism Related Genes.
Front. Public Health 10:916364.
doi: 10.3389/fpubh.2022.916364

Background: Amino acid metabolism plays a vital role in cancer biology. However, the application of amino acid metabolism in the prognosis of colon adenocarcinoma (COAD) has not yet been explored. Here, we construct an amino acid metabolism-related risk model to predict the survival outcome of COAD and improve clinical decision making.

Methods: The RNA-sequencing-based transcriptome for 524 patients with COAD from The Cancer Genome Atlas (TCGA) was selected as a training set. The integrated Gene Expression Omnibus (GEO) dataset with 1,430 colon cancer samples was used for validation. Differential expression of amino acid metabolism-related genes (AAMRGs) was identified for prognostic gene selection. Univariate cox regression analysis, LASSO-penalized Cox regression analysis, and multivariate Cox regression analysis were applied to construct a prognostic risk model. Moreover, the correlation between risk score and microsatellite instability, immunotherapy response, and drug sensitivity were analyzed.

Results: A prognostic signature was constructed based on 10 AAMRGs, including ASPG, DUOX1, GAMT, GSR, MAT1A, MTAP, PSMD12, RIMKLB, RPL3L, and RPS17. Patients with COAD were divided into high-risk and low-risk group based on the medianrisk score. Univariate and multivariate Cox regression analysis revealed that AAMRG-related signature was an independent risk factor for COAD. Moreover, COAD patients in the low-risk group were more sensitive to immunotherapy targeting PD-1 and CTLA-4.

Conclusion: Our study constructed a prognostic signature based on 10 AAMRGs, which could be used to build a novel prognosis model and identify potential drug candidates for the treatment of COAD.

Keywords: colon adenocarcinoma, amino acid metabolism, prognostic model, immune checkpoint, immune therapy

INTRODUCTION

Colon adenocarcinoma (COAD) is the most common type of colorectal cancer (1). According to the Global Cancer Observatory (GCO) (gco.iarc.fr) in 2020, there were an estimated 1.4 million new cases of colon cancer and 0.5 million deaths worldwide (2). Late diagnosis and lack of reliable biomarkers account for the poor prognosis of COAD (3). Despite many efforts, the attempt to use a single biomarker to predict the outcome of COAD has been unsuccessful (4).

Metabolic reprogramming is a common feature of tumor cells, which is crucial for rapid tumor growth and adaption to tumor microenvironment (5, 6). Apart from the well-known Warburg effect, metabolic alterations in lipid and amino acid metabolism have been observed in a number of tumors, including colorectal cancer, lung cancer, and breast cancer (7). Mounting evidence have indicated that changes in amino acid metabolism contributed to the metastasis, proliferation, angiogenesis, and drug resistance of colorectal cancer (8–11). Recent study has demonstrated that inhibition of ASCT2 (function as a glutamine transporter) exerted a great anti-tumor effect in colorectal cancer (12). Meanwhile, new insights into the metabolic signatures of tumors have revealed the potential of risk prediction model, which is based on the amino acid metabolism-related genes (AAMRG) (13, 14). In addition, amino acid metabolism plays an important role in regulating tumor immunity and targeting amino acid metabolism may help to overcome immunotherapy resistance and improve existing therapies for COAD patients. Therefore, targeting the amino acid metabolism might provide novel ideas for the diagnosis and management of COAD.

MATERIALS AND METHODS

Data Acquisition

The COAD cohort's transcriptional dataset with matching clinical information were downloaded from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/). Total 524 mRNA expression profiles including 482 COAD tissues and 42 normal colon tissues were enrolled. Three datasets with 1,430 patients with colon cancer from Gene Expression Omnibus (GEO), including GSE39582 (15), GSE29621 (16), and GSE17536 (17) were selected to verify the results of TCGA data analysis (https://www.ncbi.nlm.nih.gov/geo/). The "sva" software package in R version (4.0.2) was devoted to remove the batch effects.

Differentially Expressed AAMRGs in TCGA

Total 374 AAMRGs were extracted from the amino acid metabolism-related genes dataset (REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES), which were recorded in Molecular Signatures Database (13, 18). Then, total 327 common expressed AAMRGs in TCGA and GEO were selected. Subsequently, differentially expressed AAMRGs between the COAD and control groups were analyzed using the "limma" software package in R version (4.0.2) (based on $|\log_2 FC| > 1$ and false discovery rate < 0.05).

Construction and Validation of Prognostic Risk Score Model

The univariate Cox regression analysis was conducted for prognosis-related AAMRGs screening. The LASSO algorithm was executed to avoid overfitting the model. The multivariate Cox regression was conducted to get the optimal prognostic genes for the model. Finally, the stable AAMRGs, as the final prognosis model was constructed. We used the following equation to calculate the risk score, which was combined by regression coefficients and expression values of each AAMRG. Risk score = (index gene1 × expression of gene1) + (index gene2 × expression of gene2) + … + (index gene10 × expression of 10). All COAD patients in TCGA were divided into two subgroups (high- risk group and low-risk group) according to the median risk score. Kaplan-Meier curves were used to determine the differences in prognosis between the different groups. Finally, the first, third, fifth-year survival proportions of patients were calculated using the ROC curve. Then, the prognosis model was validated in the above merge GEO dataset.

Establishment of Nomogram Prognosis Prediction Model

We combined age, TNM stage and risk scores to plot a nomogram model using the "rms" software package in R version (4.0.2). The calibration curves were built to show the agreement between the nomogram-predicted survival probabilities and the actual survival probabilities at first-, third-, and fifth-year.

Gene Set Enrichment Analysis Between High-Risk Group and Low-Risk Group

To reveal the effect of differential expression of AAMRGs on biological pathways in COAD, Gene Set Enrichment Analysis (GSEA) was introduced to extract the potential biological function (19). Firstly, "c2.cp.kegg.v7.1.symbols.gmt" set was downloaded from Molecular Signatures Database. Secondly, "GSEA" software was applied to identify the enriched pathways in the two subgroups. The "ggplot2" software package in R version (4.0.2) was employed to visualize the top five significantly enriched biological processes in each subgroup.

Association Between Different Subgroups and Somatic Variation

Mutation data of COAD was downloaded from TCGA. The tumor mutation burden (TMB) value in each COAD patient was calculated the number of mutations in each patient. Then, differences in TMB values were analyzed between high-risk group and the low-risk group. The association between the risk score and TMB level was detected using Spearman correlation coefficient. Finally, the "maftools" software package in R version

Abbreviations: COAD, colon adenocarcinoma; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; AAMRGs, amino acid metabolism-related genes; MSI, microsatellite instability; TPM, transcripts per million; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; TCIA, The Cancer Immunome Atlas database; OS, overall survival; AUC, area under the curve; HR, Hazard ratio; TMB, tumor mutation burden; GSEA, Gene Set Enrichment Analysis.

(4.0.2) was applied to visualize the top 20 most frequently mutated genes in each group.

Correlations Between Immune Cell Infiltration Between High-Risk Group and Low-Risk Group

The CIBERSORT was performed to analyse the tumor immune microenvironment of COAD. CIBERSORT (https://cibersort.stanford.edu/) is a method of enumeration of 22 immune related cell subsets, which including naïve and memory B cells, seven types of T cell, myeloid cells, NK cells, and plasma cells (20). The "CIBERSORT" software package in R version (4.0.2) was applied to analyse the proportion of 22 immune cells between high-risk group and low-risk group. Bar plot was applied to visualize the differences in immune cells between high-risk group and low-risk group.

Differences in Immunotherapy Sensitivity Between High-Risk Group and Low-Risk Group

The immunotherapy sensitivity data of patients with COAD was obtained from The Cancer Immunome Atlas (TCIA, https://tcia.at/) which included the effectiveness score of patients with PD-1 and CTLA-4 inhibitors. Then, we detected the differences of immunotherapy sensitivity score between two subgroups.

Sensitivity Prediction of Anticancer Drugs

The prediction of the difference in drug sensitivity between high-risk group and low-risk subgroups in COAD was conducted using the "pRRophetic" software package in R version (4.0.2) based on Cancer Genome Project (CGP),which including 138 anticancer drugs (21).

Statistical Analysis

GraphPad 8.0 software and R version (4.0.2) were applied to analyse and visualize the statistical profile. The univariate Cox regression analysis, LASSO algorithm, and the multivariate Cox regression analysis were used to narrow down the number of candidate genes (22, 23). The "survival R" and "surviviner R" software package in R version (4.0.2) were used for Kaplan-Meier analysis. We compared the two groups by student's test or Wilcoxon test. The Pearson or Spearman correlation test was utilized to evaluate the correlations between variables. The $P < 0.05$ was considered statistically significant.

RESULTS AND DISCUSSION

Identifying Differential Expressions of AAMRGs Between COAD and Normal Tissues From the TCGA Dataset

The workflow of the study is displayed in **Figure 1**. Total 263 differentially expressed AAMRGs (180 upregulated genes and 83 downregulated genes) in TCGA dataset were identified. The heatmap exhibited the differentially expressed AAMRGs (**Figure 2A**).

Construction of Prognosis-Risk Signature Based on 10 AAMRGs

Fourteen AAMRGs were identified as prognostic related genes based on the univariate Cox regression analyses (**Figure 2B**). Finally, 10 prognostic related AAMRGs was screened for AAMRGs-risk signature after the LASSO analysis and the multivariate Cox regression analysis (**Figures 2C,D**). The boxplot showed the expression of 10 prognostic related AAMRGs in TCGA including ASPG (asparaginase), DUOX1 (dual oxidase 1), GAMT (guanidinoacetate N-methyltransferase), GSR (glutathione-disulfide reductase), MAT1A (methionine adenosyltransferase 1A), MTAP (methylthioadenosine phosphorylase), PSMD12 (proteasome 26S subunit, non-ATPase 12), RIMKLB (ribosomal modification protein rimK like family member B), RPL3L (ribosomal protein S17), and RPS17 (ribosomal protein L3 like) between COAD and normal samples (**Figure 2E**).

Three AAMRGs including PSMD12, MAT1A and DUOX1 were identified asprotective factors in the prognostic model. Meanwhile, seven AAMRGs including ASPG, GAMT, GSR, MTAP, RIMKLB, RPL3L and RPS17 were identified as risk factors in the prognostic model. The risk score of each COAD patient in TCGA was assessed through the equation: Risk score = $(-0.33 \times$ASPG expression$) + (0.37 \times$DUOX1 expression$) + (0.22 \times$GAMT expression$) + (-0.47 \times$GSR expression$) + (0.15 \times$MAT1A expression$) + (0.48 \times$MTAP expression$) + (-0.81 \times$PSMD12 expression$) + (0.25 \times$RIMKLB expression$) + (0.65 \times$RPL3L expression$) + (0.45 \times$RPS17 expression). Finally, the COAD patients in TCGA were divided into two groups (high-risk and low-risk group) based on the median risk score.

The **Figures 3A,C** conferred a better prognosis and longer survival time in patients with COAD in the low-risk group and worse prognosis with shorter survival time in the low-risk group (**Figure 3B**). In addition, the result of ROC curve analysis showed that the area under the curve (AUC) of first, third, fifth-year survival was 0.715, 0.750, and 0.759, which indicated a good sensitivity and specificity in predicting the prognosis of COAD based on 10 AAMRGs (**Figure 3D**). To further validate the accuracy and sensitivity of the prognosis risk signature, the above merged GEO dataset was used as an external validation dataset. Consistently, a difference in OS between high-risk and low risk group was observed ($P < 0.05$) (**Figures 3E–G**). The AUC of first, third, fifth-year survival was 0.576, 0.596, and 0.599 (**Figure 3H**).

Relationship Between the Clinicopathological Characteristics and Risk Score

Compared with the clinicopathological characteristics, the AAMRG-related prognostic risk model showed better capability in predicting one-, three-, and five-year overall survival (OS) (**Figure 4A**). Subsequently, the univariate Cox and multivariate Cox regression analyses revealed that the AAMRG-related prognostic risk model was an independent predictor of COAD prognosis (**Figure 4B**). The expression levels of the 10 screened

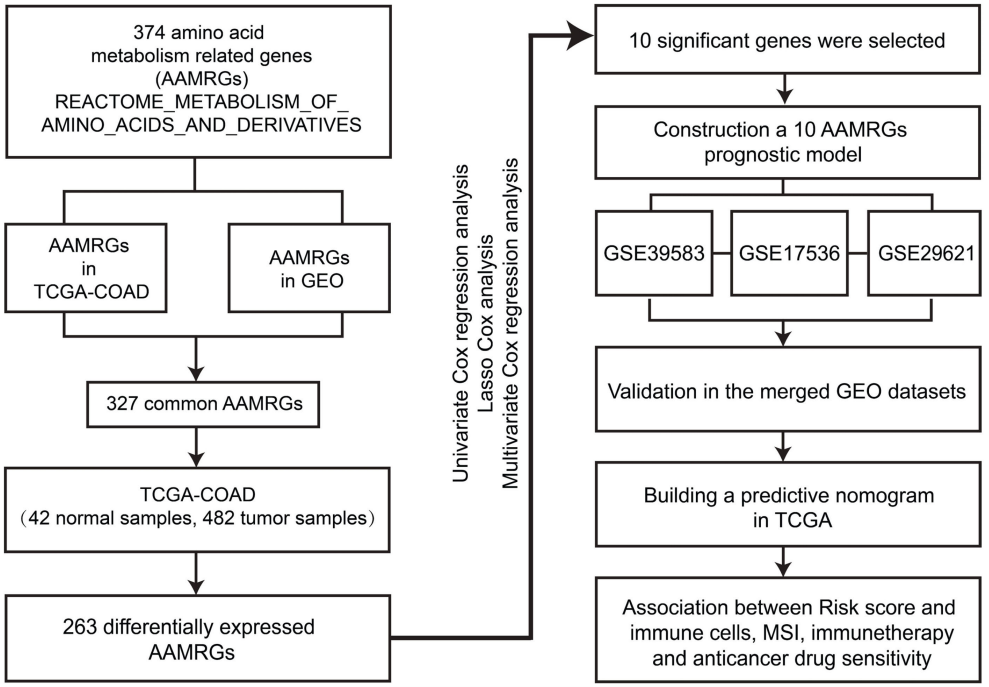


FIGURE 1 | The workflow of the study.

AAMRGs and clinicopathological characteristics between high-risk and low-risk group are depicted by heatmaps. Interestingly, there was a difference in the risk tumor stages, T, N, and M stage between high-risk and low-risk groups (**Figure 4C**). Notably, the COAD female patient in T3–T4 stage, N1–N2 stage, M1 stage, and Stage III–IV in high-risk group showed worse survival (**Supplementary Figures S1A–F**). These results indicated that the risk model may have high sensitivity and specificity for COAD patients.

Development and Evaluation of a Risk-Nomogram Based on the AAMRGs for Predicting OS in COAD Patients

The nomogram of age, stage, and risk score based on 10 AAMRGs was constructed to predict first-, third-, and fifth-year survival (**Figure 5A**). The calibration curve in **Figure 5B** demonstrated that the prediction and actuality of fifth-year survival values were in good agreement.

GSEA With the Prognostic Risk Signature

The GSEA was conducted between the high-risk and low-risk group based on the prognosis-risk scores. As displayed in **Supplementary Figure S2A**, the melanoma related signal pathway, ECM receptor interaction, WNT signaling pathway, mTOR signaling pathway and *TGF-β* signaling pathway might be positively correlated with the higher risk scores in COAD patients. In addition, the porphyrin and chlorophyll metabolism, proteasome, pentose and glucuronate interconversions, citrate cycle, TCA cycle, and drug metabolism related signal pathway were negatively correlated with COAD patients in high-risk group.

The Correlations Between Tumor Microenvironment Cell Infiltration Characteristics and Risk Score

The CIBERSORT results showed that activated NK cells, eosinophils, and neutrophils were more abundant in patients in the low-risk group ($P < 0.05$) (**Figure 6**). On the other hand, the infiltration of monocytes increased significantly in the high-risk group.

Correlations Between the Risk Score Model and Somatic Variants

The TMB levels were calculated between high-risk and low-risk groups. However, there was no significant difference in TMB between the two groups (**Supplementary Figure S2B**). However, the somatic mutations of TTN, SYNE1, PIK3CA, MUC16, FAT4, ZFHX4, RYR2, OBSCN, DNAH5, PLCO were more common in the low-risk group, whereas the mutation frequency of APC, TP53, and KRAS mutations was higher in the high-risk group (**Supplementary Figures S2C,D**).

Risk Score Predicts Resistance to Immunotherapy

The COAD Patients from TCIA database were divided into three groups based on the MSI status: high microsatellite instability (MSI-H), microsatellite-stable (MSS), and low microsatellite instability (MSI-L). As shown in **Figure 7A**, the proportion of

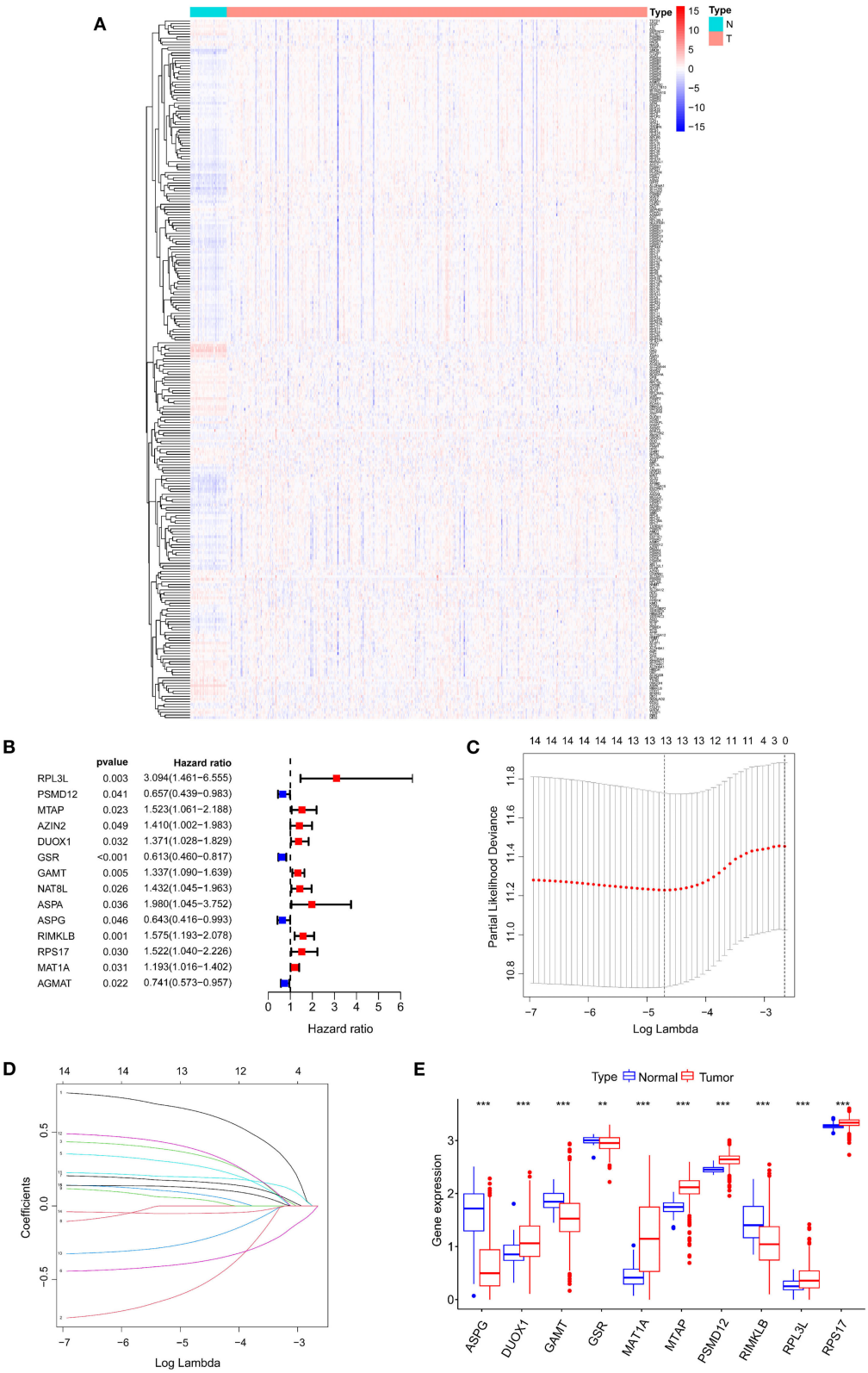


FIGURE 2 | Identifying candidate genes associated with the prognosis of COAD patients. **(A)** Differential expression heatmap of amino acid metabolism-related genes in COAD and normal tissues from the TCGA. **(B)** Univariate Cox regression analysis of AAMRGs. **(C)** Turning optimal parameter (*lambda*) screening in the LASSO model. **(D)** LASSO coefficient profiles of the prognostic genes. **(E)** Box plot of mRNA expression of AAMRGs. $^{**}P < 0.01$ and $^{***}P < 0.001$.

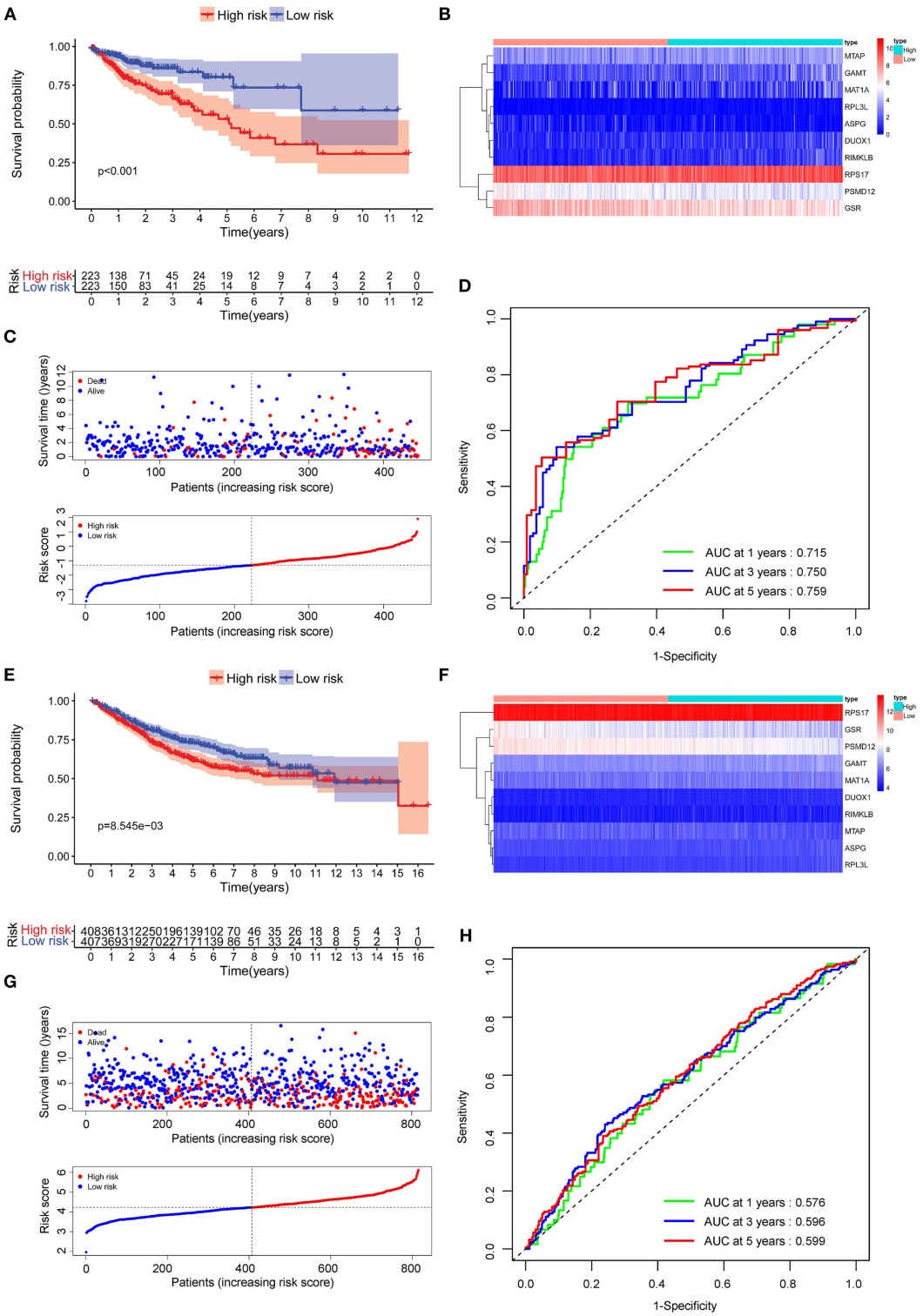


FIGURE 3 | Evaluation of the prognostic performance of the AAMRGs signature in the TCGA dataset and GEO dataset. **(A)** The Kaplan-Meier survival curves of the AAMRGs signature in TCGA dataset. Patients from the TCGA dataset were stratified into two groups according to the median risk scores. **(B,C)** The distribution of risk score, survival outcomes, and the expression pattern of 10 AAMRGs in the TCGA dataset. **(D)** The ROC curves for predicting the 1-, 3-, and 5-year OS in the TCGA dataset. **(E)** The Kaplan-Meier survival curves of the AAMRGs signature in GEO dataset. Patients from the GEO dataset were stratified into two groups according to the median risk score. **(F,G)** The distribution of risk score, survival outcomes, and the expression pattern of nine AAMRGs in the GEO dataset. **(H)** The ROC curves for predicting the 1-, 3-, and 5-year OS in the GEO dataset.

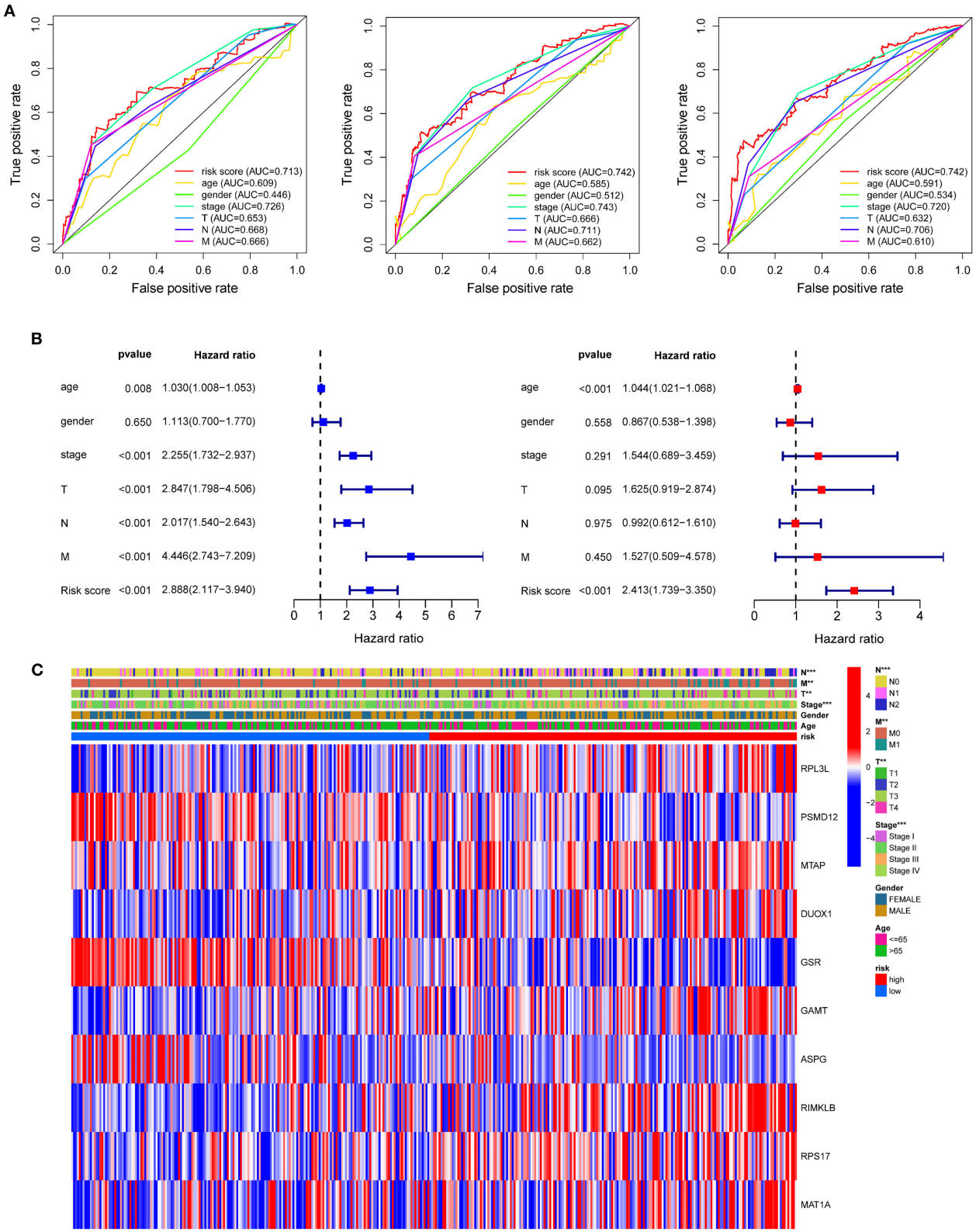


FIGURE 4 | The predictive efficacy of the risk model and the relationship between risk score and clinical outcome, pathological characteristics, and prognostic value of COAD. **(A)** ROC curve to evaluate the predictive efficacy of the risk model. **(B)** Univariate Cox regression analysis and multivariate Cox regression analysis of clinicopathological characteristics and risk score. **(C)** The relationship of clinicopathological characteristics and risk scores between high- and low-risk groups from TCGA dataset. $^{**}P < 0.01$ and $^{***}P < 0.001$.

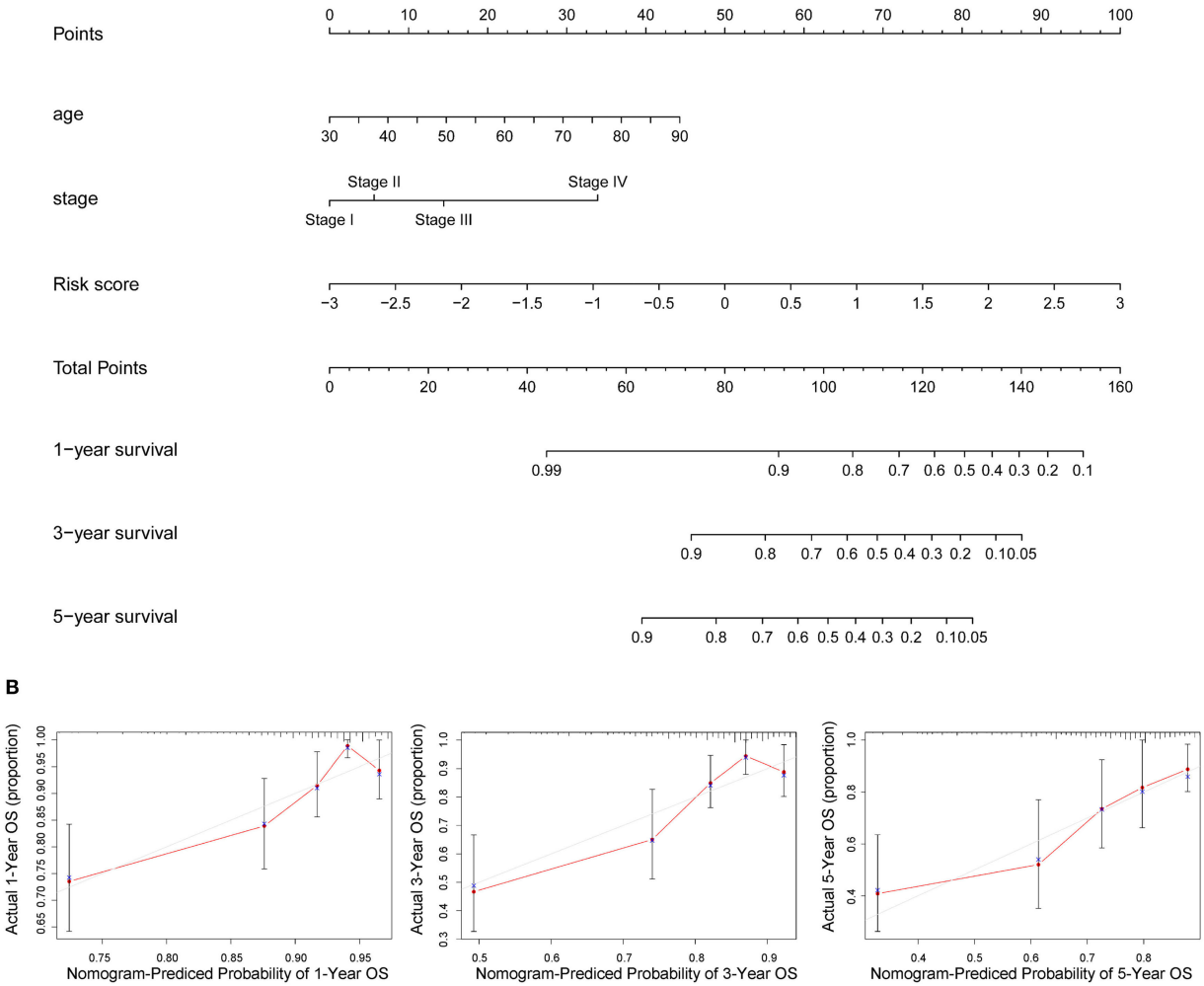


FIGURE 5 | Prognostic nomogram incorporating the risk score model and clinicopathological characteristics. **(A)** The nomogram of age, stage, and risk score for predicting 1-, 3-, and 5-year survival. **(B)** The 1-, 3-, and 5-year calibration curves of TCGA dataset.

MSI-H patients in low-risk group (26%) was higher than the high-risk group (12%). The results showed that patients in the MSI-L and MSS groups had higher risk scores, compared with MSI-H group ($P < 0.05$) (**Figure 7B**). This indicated that patients with lower risk scores were more sensitive to immunotherapy. Notably, the COAD patients in low-risk group were more sensitive to the combination of PD-1 and CTLA-4 inhibitors ($P < 0.05$) or CTLA-4 inhibitors alone ($P < 0.05$) than in the high-risk group (**Figures 7C,D**). However, there was no difference observed in the sensitivity PD-1 inhibitors alone between the two subgroups (**Figures 7E,F**). These data suggest that the risk score of COAD patients may affect the immunotherapy selection in COAD patients.

Results of Risk Score Model and Drug Sensitivity

Four drugs including imatinib ($P = 6.8e{-}09$, **Figure 8A**), midostaurin ($P = 1.2e{-}06$, **Figure 8B**), pazopanib ($P = 3.7e{-}04$, **Figure 8C**), and elesclomol ($P = 8.2e{-}03$, **Figure 8D**) were identified with lower IC50 level in high-risk group of COAD patients, which are suggestive of better efficacy. Besides, we found that COAD patients in the low-risk group were more sensitive to drugs including paclitaxel, metformin, rapamycin, bortezomib, sorafenib, and gemcitabine (**Supplementary Figures S3A–F**).

DISCUSSION

Colorectal cancer is one of the most common types of malignant tumors. The majority of COAD patients in high-risk stage II and stage III were recommended to receive surgery combined with adjuvant chemotherapy to reduce the risk of recurrence. However, approximately half of the patients in the early stage received radical surgery developed recurrence and metastasis (24). Therefore, there is still an urgent need for constructing a prognostic model that

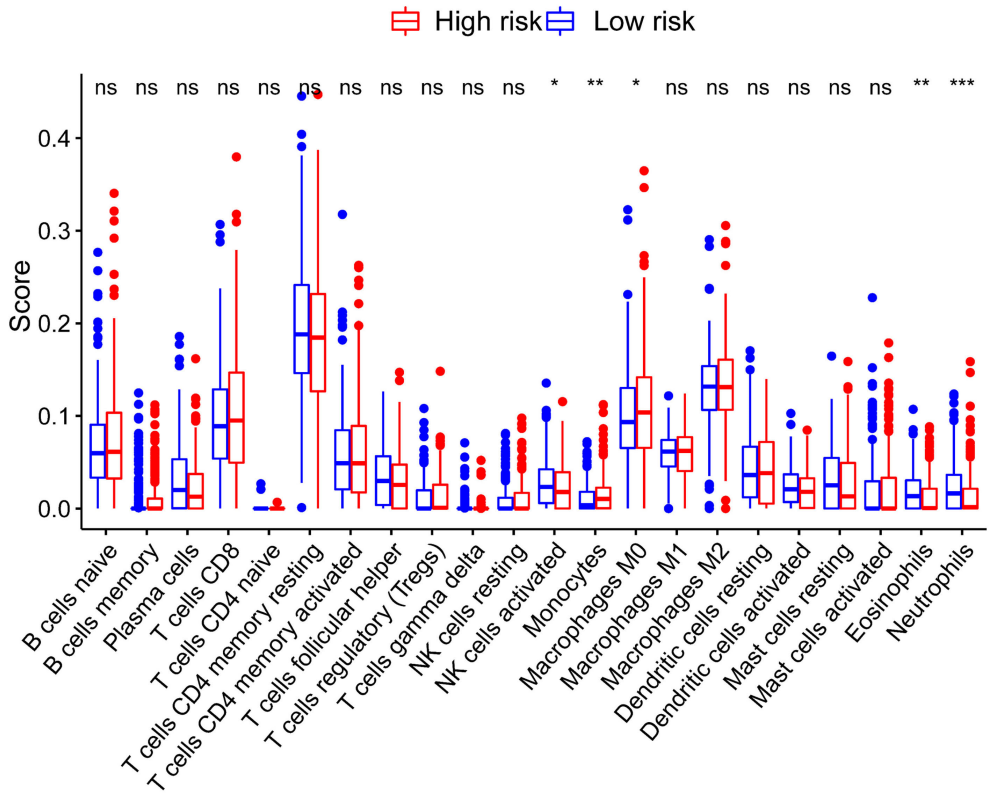


FIGURE 6 | Correlations between the risk score model and tumor-infiltrating immune cells. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

provides personalized prognosis and precision medicine for COAD patients.

Characteristic metabolic changes in malignant cells were observed including abnormalities in amino acid metabolism, increased fatty acid synthesis and glucose uptake (25). The reprogramming of amino acids played an essential role in the tumorigenesis and tumor progression. The abnormalities in amino acid metabolism also deeply reshaped the tumor microenvironment, especially the function of immune cells (26). Several studies have demonstrated that amino acid metabolism could be a therapeutic target in multiple solid tumors (27–33). Therefore, we constructed and validated a predictive model based on AAMRGs to predict the prognosis of COAD patients.

In this study, we analyzed the differentially expressed genes related to amino acid metabolism in TCGA by performing the univariate Cox regression analysis, LASSO algorithm, and multivariate Cox regression analysis. Ten AAMRGs (ASPG, DUOX1, GAMT, GSR, MAT1A, MTAP, PSMD12, RIMKLB, RPL3L, and RPS17) were screened to construct a prognosis risk model for prediction. GSR was downregulated in COAD and inhibited the metastasis of colon cancer cell (34). In addition, GSR plays an important role in the conversion of GSSG to GSH in COAD (35). MTAP was upregulated in COAD and could accelerate the proliferation, invasion and migration of COAD (36, 37). However, the functional role of PSMD12, MAT1A, ASPG, GAMT, RIMKLB, RPL3L, and RPS17 in COAD remains unknown. The accuracy and sensitivity of the model were further validated with a merged GEO dataset. Our results indicated that high-risk is linked to worse prognosis.

To explore the correlations between immune cell infiltration and risk score, we use CIBERSORT algorithm and estimate the difference in immune infiltration between two subgroups. We found that the levels of monocytes, activated NK cells, eosinophils, and neutrophils were significantly increased in the low-risk group. These results suggested that the amino acid metabolism-related gene signature may affect the infiltration of immune cells and potentially the response of immunotherapy.

It was known that TMB is associated with the production of neoantigens and the response of immunotherapy in various tumors (38). There was only a small population of COAD patients who benefited from immunotherapy (39–42). Currently, TMB and MSI are the best predictors of the therapeutic effects of immune checkpoint inhibitors (ICIs) in COAD patients (43, 44). We further analyzed the relationship between the risk score and MSI. The low-risk group was found to have higher MSI level and increased sensitivity to immunotherapy. Data in the TCIA database showed that patients with low risk score might show greater sensitivity to the combination of PD-1 and CTLA-4 inhibitors and

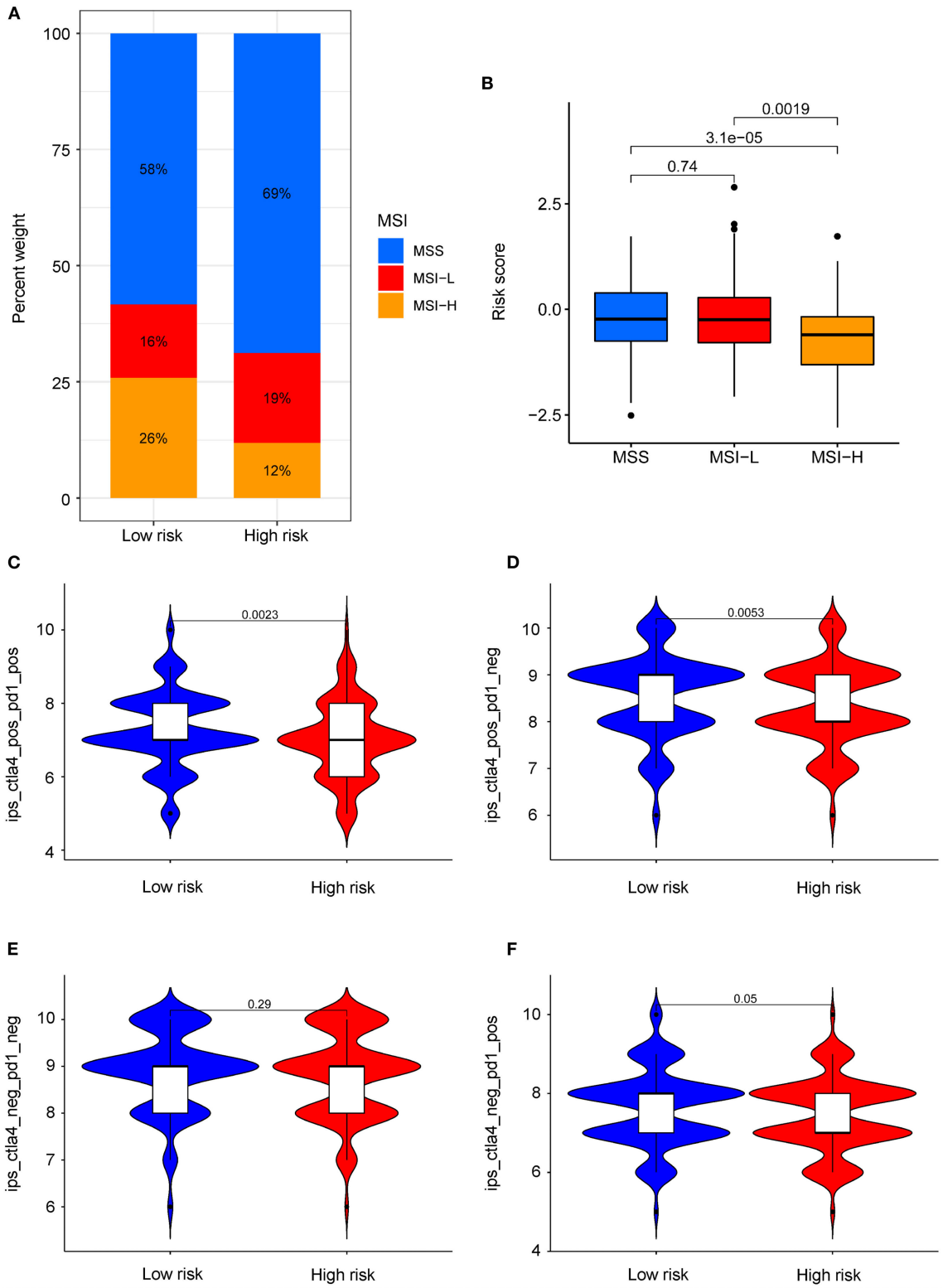


FIGURE 7 | Role of risk score in predicting MSI and immunotherapy benefits. **(A)** The proportion of different MSI levels in the subgroups with high and low risk score. **(B)** Differences in risk score among groups with different MSI levels. **(C–F)** Sensitivity of patients with high and low risk score subgroups to four treatments. **(C)** PD-1 inhibitor in combination with CTLA-4 inhibitor. **(D)** CTLA-4 inhibitor alone. **(E)** Without immune checkpoint inhibitors. **(F)** PD-1 inhibitor alone.

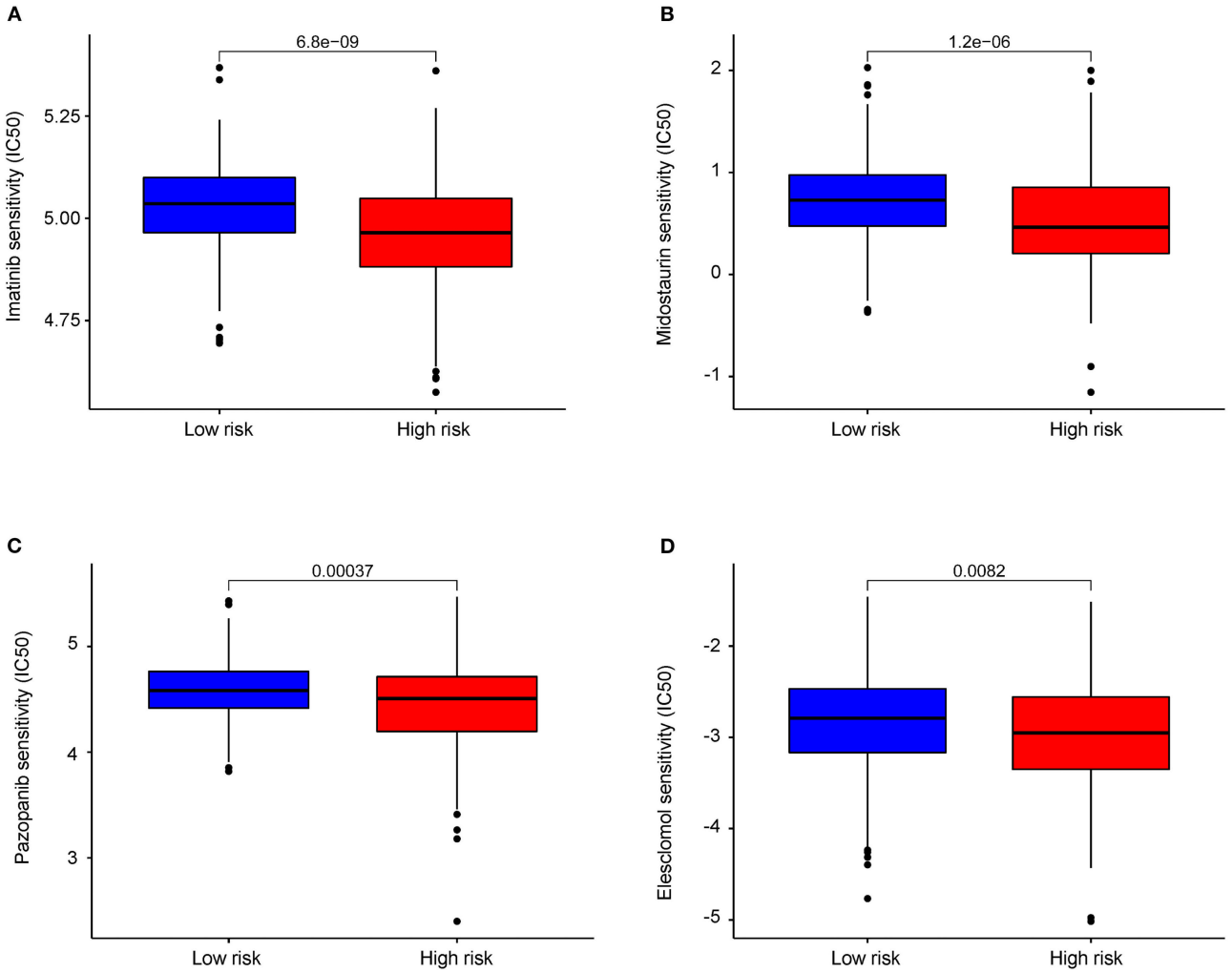


FIGURE 8 | Differential chemotherapeutic response based on IC50 between the high- and low-AAMRGs-risk groups. **(A–D)** The half-maximal inhibitory concentration (IC50) of 4 chemotherapeutic agents (imatinib, midostaurin, pazopanib and elesclomol).

CTLA-4 inhibitors alone. In conclusion, dual CTLA-4/PD-1 blockade might be considered as suitable drug for patients with low-risk scores.

The combination of traditional chemotherapy drugs and targeted therapy has been widely used in the treatment of advanced colon cancer. We found that patients in the high-risk group had a higher sensitivity to elesclomol, midostaurin, pazopanib, and imatinib than in the low-risk group. Elesclomol, a reactive oxygen species (ROS) inducer, plays an important role in mediating cuproptosis (45). Pazopanib is a multi-targeted receptor tyrosine kinase inhibitor that selectively restraining the autophosphorylation of receptors such as VEGFR-2, Kit, and PDGFR-β in renal cell carcinoma (46). In our study, we may provide a new complementary for the treatment of advanced colon cancer. However, the results of this drug screening still need further validation in clinical trials.

CONCLUSIONS

In summary, we constructed a prognostic signature based on 10 AAMRGs with strong predictive value. This study paved the way for personalized treatment of COAD patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE39582, GSE29621, and GSE17536.

AUTHOR CONTRIBUTIONS

YR and WY conceived and designed the study. YR and SF performed the experiments and analyzed the data. YR and

SH wrote the manuscript. WY contributed to manuscript revision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Guangdong Natural Science Funds for Distinguished Young Scholar (No. 2017A030306030 to WY).

ACKNOWLEDGMENTS

The authors express their gratitude to the editor and the reviewers for their useful comments and suggestions that helped improve the quality of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.916364/full#supplementary-material

Supplementary Figure S1 | Kaplan-Meier survival subgroup analysis according to the signature stratified by clinical characteristics. **(A)** Age ≤ 70 years and age > 70 years. **(B)** Female and male. **(C)** T1-2 and T3-4. **(D)** N0 and N1-2. **(E)** M0 and M1. **(F)** Stage I–II and stage III–IV.

Supplementary Figure S2 | Correlation between risk score and GSEA, somatic variation. **(A)** Enriched gene sets annotated by the KEGG collection between the high- and low-AAMRGs-risk groups in the cohort. **(B)** TMB levels between the high- and low-risk groups. **(C)** Correlation analysis between risk score and mutation load. **(D,E)** The mutation rates of reported prognostic-related genes in low- and high-risk groups.

Supplementary Figure S3 | Differential chemotherapeutic response based on IC50 available between the high- and low-risk groups. **(A–F)** The half-maximal inhibitory concentration (IC50) of six chemotherapeutic agents (Paclitaxel, Metformin, Rapamycin, Bortezomib, Sorafenib, Gemcitabine).

REFERENCES

1. Mutch MG. Molecular profiling and risk stratification of adenocarcinoma of the colon. *J Surg Oncol.* (2007) 96:693–703. doi: 10.1002/jso.20915
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
3. Viscaino M, Torres Bustos J, Muñoz P, Auat Cheein C, Cheein FA. Artificial intelligence for the early detection of colorectal cancer: a comprehensive review of its advantages and misconceptions. *World J Gastroenterol.* (2021) 27:6399–414. doi: 10.3748/wjg.v27.i38.6399
4. Chang YC, Chan MH, Li CH, Fang CY, Hsiao M, Chen CL. Exosomal components and modulators in colorectal cancer: novel diagnosis and prognosis biomarkers. *Biomedicines.* (2021) 9:931. doi: 10.3390/biomedicines9080931
5. Martínez-Reyes I, Chandel NS. Cancer metabolism: looking forward. *Nat Rev Cancer.* (2021) 21:669–80. doi: 10.1038/s41568-021-00378-6
6. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. *Science.* (2020) 368:eaaw5473. doi: 10.1126/science.aaw5473
7. Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. *Cell Mol Life Sci.* (2016) 73:377–92. doi: 10.1007/s00018-015-2070-4
8. Prosén S, Tina E, Sneckenborg AH, Loinder C, Seifert O, Lindberg M, et al. Increased expression of LAT1 in basal cell carcinoma: implications for tumour cell survival. *Clin Exp Dermatol.* (2022) 47:910–7. doi: 10.1111/ced.15038
9. Hua Q, Zhang B, Xu G, Wang L, Wang H, Lin Z, et al. CEMIP, a novel adaptor protein of OGT, promotes colorectal cancer metastasis through glutamine metabolic reprogramming *via* reciprocal regulation of β-catenin. *Oncogene.* (2021) 40:6443–55. doi: 10.1038/s41388-021-02023-w
10. Quan L, Ohgaki R, Hara S, Okuda S, Wei L, Okanishi H, et al. Amino acid transporter LAT1 in tumor-associated vascular endothelium promotes angiogenesis by regulating cell proliferation and VEGF-A-dependent mTORC1 activation. *J Exp Clin Cancer Res.* (2020) 39:266. doi: 10.1186/s13046-020-01762-0
11. Pranzini E, Pardella E, Paoli P, Fendt SM, Taddei ML. Metabolic reprogramming in anticancer drug resistance: a focus on amino acids. *Trends Cancer.* (2021) 7:682–99. doi: 10.1016/j.trecan.2021.02.004
12. Schulte ML, Fu A, Zhao P, Li J, Geng L, Smith ST, et al. Pharmacological blockade of ASCT2-dependent glutamine transport leads to antitumor efficacy in preclinical models. *Nat Med.* (2018) 24:194–202. doi: 10.1038/nm.4464
13. Liu YQ, Chai RC, Wang YZ, Wang Z, Liu X, Wu F, et al. Amino acid metabolism-related gene expression-based risk signature can better predict overall survival for glioma. *Cancer Sci.* (2019) 110:321–33. doi: 10.1111/cas.13878
14. Zhao Y, Zhang J, Wang S, Jiang Q, Xu K. Identification and validation of a nine-gene amino acid metabolism-related risk signature in HCC. *Front Cell Dev Biol.* (2021) 9:731790. doi: 10.3389/fcell.2021.731790
15. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* (2013) 10:e1001453. doi: 10.1371/journal.pmed.1001453
16. Chen DT, Hernandez JM, Shibata D, McCarthy SM, Humphries LA, Clark W, et al. Complementary strand microRNAs mediate acquisition of metastatic potential in colonic adenocarcinoma. *J Gastrointest Surg.* (2012) 16:905–12. doi: 10.1007/s11605-011-1815-0
17. Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B, Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology.* (2010) 138:958–68. doi: 10.1053/j.gastro.2009.11.005
18. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
20. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* (2015) 12:453–7. doi: 10.1038/nmeth.3337
21. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE.* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468
22. Lakshmanna K, Khare N. Constraint-based measures for DNA sequence mining using group search optimization algorithm. *Int J Intell Eng Syst.* (2016) 9:91–100. doi: 10.22266/ijies2016.0930.09
23. Lakshmanna K, Khare N. Mining DNA sequence patterns with constraints using hybridization of firefly and group search optimization. *J Intell Syst.* (2017) 27:349–62. doi: 10.1515/jisys-2016-0111
24. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin.* (2016) 66:271–89. doi: 10.3322/caac.21349
25. Schiliro C, Firestein BL. Mechanisms of metabolic reprogramming in cancer cells supporting enhanced growth and proliferation. *Cells.* (2021) 10:1059. doi: 10.3390/cells10051056

26. Commisso C, Davidson SM, Soydaner-Azeloglu RG, Parker SJ, Kamphorst JJ, Hackett S, et al. Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature.* (2013) 497:633–7. doi: 10.1038/nature12138
27. Endicott M, Jones M, Hull J. Amino acid metabolism as a therapeutic target in cancer: a review. *Amino Acids.* (2021) 53:1169–79. doi: 10.1007/s00726-021-03052-1
28. Glazer ES, Piccirillo M, Albino V, Di Giacomo R, Palaia R, Mastro AA, et al. Phase II study of pegylated arginine deiminase for nonresectable and metastatic hepatocellular carcinoma. *J Clin Oncol.* (2010) 28:2220–6. doi: 10.1200/JCO.2009.26.7765
29. Ascierto PA, Scala S, Castello G, Daponte A, Simeone E, Ottaiano A, et al. Pegylated arginine deiminase treatment of patients with metastatic melanoma: results from phase I and II studies. *J Clin Oncol.* (2005) 23:7660–8. doi: 10.1200/JCO.2005.02.0933
30. Kelly MP, Jungbluth AA, Wu BW, Bomalaski J, Old LJ, Ritter G. Arginine deiminase PEG20 inhibits growth of small cell lung cancers lacking expression of argininosuccinate synthetase. *Br J Cancer.* (2012) 106:324–32. doi: 10.1038/bjc.2011.524
31. Miraki-Moud F, Ghazaly E, Ariza-McNaughton L, Hodby KA, Clear A, Anjos-Afonso F, et al. Arginine deprivation using pegylated arginine deiminase has activity against primary acute myeloid leukemia cells *in vivo*. *Blood.* (2015) 125:4060–8. doi: 10.1182/blood-2014-10-608133
32. Han Q, Tan Y, Hoffman RM. Oral dosing of recombinant methioninase is associated with a 70% drop in PSA in a patient with bone-metastatic prostate cancer and 50% reduction in circulating methionine in a high-stage ovarian cancer patient. *Anticancer Res.* (2020) 40:2813–19. doi: 10.21873/anticanres.14254
33. Blackburn PR, Gass JM, Vairo FPE, Farnham KM, Atwal HK, Macklin S, et al. Maple syrup urine disease: mechanisms and management. *Appl Clin Genet.* (2017) 10:57–66. doi: 10.2147/TACG.S125962
34. Gu L, Liu Y, Jiang C, Sun L, Zhou H. Identification and clinical validation of metastasis-associated biomarkers based on large-scale samples in colon-adenocarcinoma. *Pharmacol Res.* (2020) 160:105087. doi: 10.1016/j.phrs.2020.105087
35. Zhao M, Liu Q, Gong Y, Xu X, Zhang C, Liu X, et al. GSH-dependent antioxidant defense contributes to the acclimation of colon cancer cells to acidic microenvironment. *Cell Cycle.* (2016) 15:1125–33. doi: 10.1080/15384101.2016.1158374
36. Bataille F, Rogler G, Modes K, Poser I, Schuierer M, Dietmaier W, et al. Strong expression of methylthioadenosine phosphorylase (MTAP) in human colon carcinoma cells is regulated by TCF1/[beta]-catenin. *Lab Invest.* (2005) 85:124–36. doi: 10.1038/labinvest.3700192
37. Zhong Y, Lu K, Zhu S, Li W, Sun S. Characterization of methylthioadenosin phosphorylase (MTAP) expression in colorectal cancer. *Artif Cells Nanomed Biotechnol.* (2018) 46:2082–7. doi: 10.1080/21691401.2017.1408122
38. Song W, Shen L, Wang Y, Liu Q, Goodwin TJ Li J, et al. Synergistic and low adverse effect cancer immunotherapy by immunogenic chemotherapy and locally expressed PD-L1 trap. *Nat Commun.* (2018) 9:2237. doi: 10.1038/s41467-018-04605-x
39. Tolba MF. Revolutionizing the landscape of colorectal cancer treatment: the potential role of immune checkpoint inhibitors. *Int J Cancer.* (2020) 147:2996–3006. doi: 10.1002/ijc.33056
40. Munari E, Mariotti FR, Quatrini L, Bertoglio P, Tumino N, Vacca P, et al. PD-1/PD-L1 in Cancer: pathophysiological, diagnostic and therapeutic aspects. *Int J Mol Sci.* (2021) 22:5123. doi: 10.3390/ijms22105123
41. Zhang H, Dai Z, Wu W, Wang Z, Zhang N, Zhang L, et al. Regulatory mechanisms of immune checkpoints PD-L1 and CTLA-4 in cancer. *J Exp Clin Cancer Res.* (2021) 40:184. doi: 10.1186/s13046-021-01987-7
42. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet.* (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8
43. Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol.* (2019) 16:361–75. doi: 10.1038/s41575-019-0126-x
44. Gorzo A, Galos D, Volovat SR, Lungulescu CV, Burz C, Sur D. Landscape of immunotherapy options for colorectal cancer: current knowledge and future perspectives beyond immune checkpoint blockade. *Life.* (2022) 12:229. doi: 10.3390/life12020229
45. Hersey P, Bastholt L, Chiarion-Sileni V, Cinat G, Dummer R, Eggermont AM, et al. Small molecules and targeted therapies in distant metastatic disease. *Ann Oncol.* (2009) 20:vi35–40. doi: 10.1093/annonc/mdp254
46. Keisner SV, Shah SR. Pazopanib: the newest tyrosine kinase inhibitor for the treatment of advanced or metastatic renal cell carcinoma. *Drugs.* (2011) 71:443–54. doi: 10.2165/11588960-000000000-00000

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ren, He, Feng and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fpubh.2022.899027

A Prediction Model for Peak Expiratory Flow Derived From Venous Blood Biomarkers and Clinical Factors in Amyotrophic Lateral Sclerosis

Xianghua He[1,2], *Jiaming Feng*[3], *Xue Cong*[4], *Hongyan Huang*[1], *Quanzhen Zhao*[1], *Qiuyan Shen*[1], *Fang Xu*[1] *and Yanming Xu*[1]*

[1] Department of Neurology, West China Hospital, Sichuan University, Chengdu, China, [2] Department of Neurology, Jiangbin Hospital, Nanning, China, [3] West China Clinical Medical College, Sichuan University, Chengdu, China, [4] West China School of Public Health, Sichuan University, Chengdu, China

Although peripheral venous blood biomarkers are related to respiratory function in Amyotrophic lateral sclerosis (ALS) patients, there are still few prediction models that predict pulmonary function. This study aimed to investigate the venous blood biomarkers associated with respiratory function in patients with ALS from southwest China and to create prediction models based on those clinical biomarkers using logistic regression. A total of 319 patients with ALS from the retrospective cohort and 97 patients with ALS from the prospective cohort were enrolled in this study. A multivariable prediction model for the correlation between peak expiratory flow (PEF) and hematologic, biochemical laboratory parameters, and clinical factors in patients with ALS was created. Along with female patients, bulbar-onset, lower body mass index (BMI), later age of onset, lower level of creatinine, uric acid, triglyceride, and a higher level of high-density lipoprotein cholesterol (HDL_C) were related to reduced PEF. The area under the receiver operating characteristics (ROC) curve is.802 for the test set and.775 for the validation set. The study constructed a multivariable prediction model for PEF in patients with ALS. The results can be helpful for clinical practice to predict respiratory impairment.

Keywords: Amyotrophic lateral sclerosis, peak expiratory flow, prediction model, respiratory function, venous blood parameters

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Md Habibur Rahman, Islamic University, Bangladesh
Srinivasa Rao Namburi, Bapatla Engineering College, India

***Correspondence:**
Yanming Xu
neuroxym999@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *18 March 2022*
Accepted: *29 April 2022*
Published: *27 May 2022*

Citation:
He X, Feng J, Cong X, Huang H, Zhao Q, Shen Q, Xu F and Xu Y (2022) A Prediction Model for Peak Expiratory Flow Derived From Venous Blood Biomarkers and Clinical Factors in Amyotrophic Lateral Sclerosis. Front. Public Health 10:899027. doi: 10.3389/fpubh.2022.899027

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disease clinically characterized by loss of bulbar and limb function (1). Recent studies have shown that the average onset age of ALS is 51–66 years (2). Men have a higher risk of developing ALS than women. Studies reported a male to female ratio of 1:2 (2). Patients with ALS usually die from respiratory failure 3–5 years after the symptom onset (3).

To monitor the health status of patients with ALS, close clinical and functional follow-up with particular care for respiratory monitoring was recommended for patients with ALS (4).

Peak expiratory flow (PEF) can assess the strength of respiratory muscles and can evaluate the efficacy of clearing the bronchial tree (5). The study reported that 85% of patients with ALS would develop dysphagia and impaired secretion management, leading to impairment of mucociliary clearing and airway protection (6), which resulted in malnutrition, dehydration, and aspiration pneumonia (7). PEF can be used as a tool for assessing cough ability for patients who underwent lung surgery (8), pulmonary function assessment before lung surgery (9, 10), and airway clearance in neurodegenerative disease (11). Recent studies reported that respiratory impairments in patients with ALS demonstrated reduced peak expiratory flow (PEF) rates compared to healthy controls (12, 13), implying that PEF may be a good screen tool for identifying patients with ALS at risk for aspiration (13). Moreover, the authors found that PEF was an independent factor for survival (14) and positively correlated with the revised functional rating scale score (ALSFRS-R), indicating a predictive value for survival and disease severity in patients with ALS (15).

For patients in the final stages of ALS, the pulmonary function tests were hardly carried out due to their inability to perform the tests. Thus, it is necessary to explore more easily accessible and low-cost parameters for the prediction of respiratory function. Venous blood parameters have been related to respiratory impairment and disease severity in patients with ALS in several studies (16–20). To the best of our knowledge, no studies of venous blood biomarkers associated with PEF in patients with ALS were seen.

We aimed to develop and validate a predictive model based on a new panel of eleven variables analyzed in venous blood parameters of patients with ALS, along with their clinical characteristics. The multivariable logistic regression model created in this study can predict the impairment of PEF in patients with ALS.

MATERIALS AND METHODS

Patients

This is a retrospective and prospective study. Patients from a large, regional, and referral ALS clinic in Southwest China, were enrolled in the study. In the retrospective cohort, 319 patients with ALS were registered from 1 January 2015 to 30 December 2020. In the prospective cohort, 97 participants from 1 March 2017 to 30 October 2021 were enrolled. All the participants fulfilled the revised El Escorial criteria for probable or definite ALS (21, 22). We excluded patients with other medical or neurological diseases and patients with missing baseline hematologic and biochemical values. Patients with manifestations of neoplastic disorders were also excluded. This study was approved by the Ethical Committee of West China Hospital of Sichuan University.

Clinical Variables

In this study, the relevant demographic and clinical data were collected, including age, age at ALS onset and diagnosis, sex, onset body region, and blood pressure. BMI was calculated at

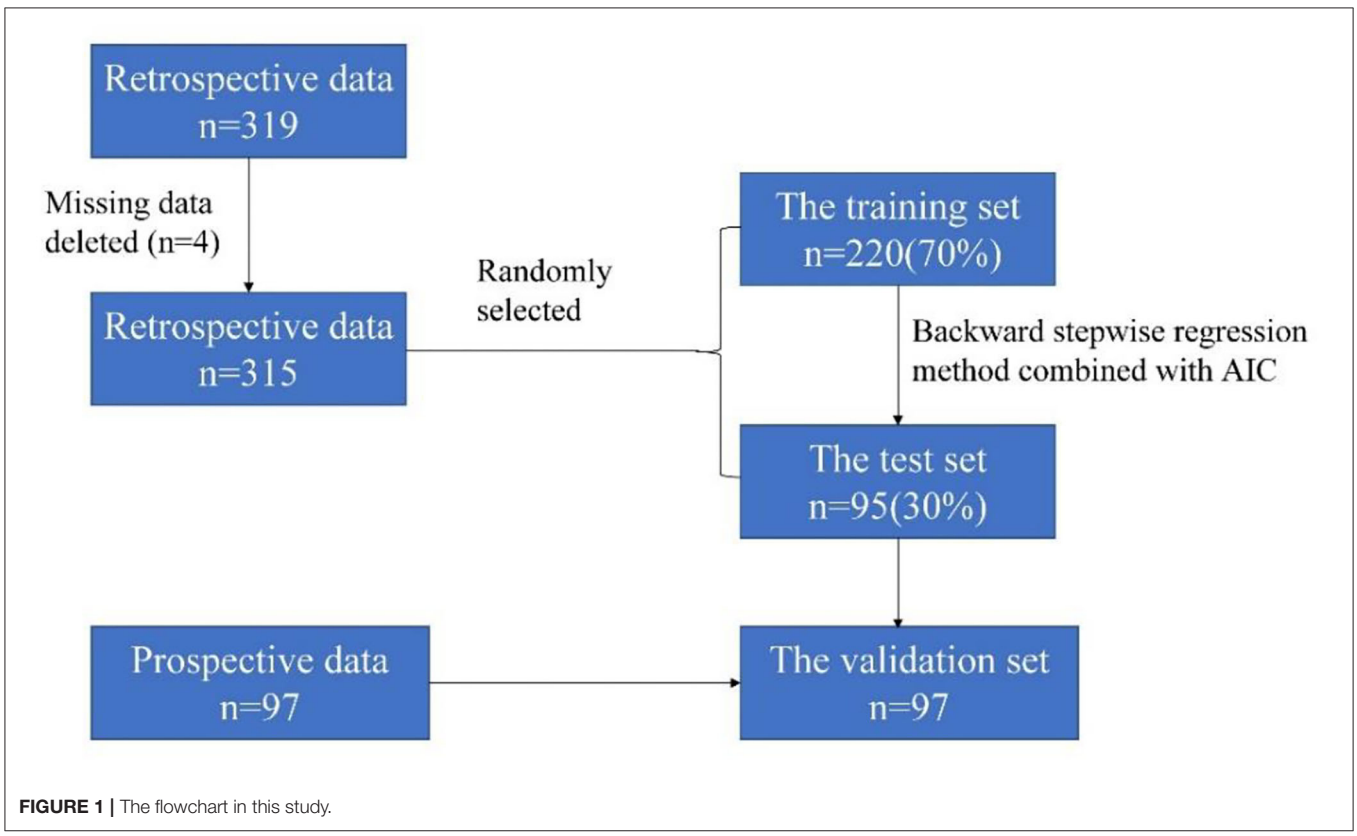


FIGURE 1 | The flowchart in this study.

TABLE 1 | Clinicodemographic data of Amyotrophic lateral sclerosis patients stratified by whether their PEF is ≥5.50 L/s.

	The retrospective cohort			The prospective cohort		
Variable	**PEF ≥ 5.5 (*n* = 208)**	**PEF < 5.5 (*n* = 111)**	***P***	**PEF ≥ 5.5 (*n* = 63)**	**PEF < 5.5 (*n* = 34)**	***P***
Gender						
Male	173 (83.17%)	57 (51.35%)	<0.001	43 (68.25%)	16 (47.06%)	0.041
female	35 (16.83%)	54 (48.65%)		20 (31.75%)	18 (52.94%)	
age	52.59 ± 10.39	59.64 ± 10.40	<0.001	53.37 ± 10.57	61.4 4± 11.83	0.001
onset age	51.44 ± 10.38	58.26 ± 10.91	<0.001	52.35 ± 10.39	60.47 ± 11.90	0.001
Hypertension						
SBP	131.18 ± 16.88	133.59 ± 16.74	0.229	130.03 ± 21.43	129.18 ± 17.80	0.843
DBP	88.31 ± 13.24	85.79 ± 11.96	0.098	85.30 ± 13.04	84.65 ± 13.25	0.655
weight	60.00 (13.00)	55.00 (12.00)	<0.001	60.92 ± 8.97	52.88 ± 7.35	<0.001
height	1.63 ± 0.075	1.57 ± 0.074	<0.001	1.62 ± 0.07	1.58 ± 0.08	0.012
BMI	22.93 ± 2.98	22.16 ± 3.07	0.03	23.21 ± 2.70	21.28 ± 2.96	0.002
Site of onset						
limb onset	179 (86.06%)	80 (72.73%)	0.007	49 (77.78%)	26 (76.47%)	0.883
bulbar onset	23 (11.06%)	27 (24.54%)		14 (22.22%)	8 (23.53%)	
other	6 (2.88%)	3 (2.73%)		0	0	
Diagnostic level						
probable	102 (49.04%)	53 (47.75%)	0.826	46 (73.02%)	24 (70.59%)	0.799
definite	106 (50.96%)	58 (52.25%)		17 (26.98%)	10 (29.41%)	
PLT	177.50 (79.00)	195.00 (78.00)	0.052	167.00 (80.00)	188.50 (54.50)	0.084
NEUT#	3.39 (1.51)	3.41 (1.67)	0.644	3.21 ±1.06	3.44 ± 1.03	0.552
LYMPH#	1.79 (0.62)	1.62 (0.79)	0.035	1.60 (0.79)	1.39 (0.87)	0.165
PLR	101.26 (52.94)	114.52 (62.55)	0.002	94.04 (61.65)	128.60 (66.52)	0.027
MONO#	0.41 (0.16)	0.38 (0.19)	0.283	0.41 ± 0.12	0.38 ± 0.14	0.632
EO#	0.12 (0.12)	0.12 (0.12)	0.508	0.14 (0.16)	0.14 (0.15)	0.797
BASO#	0.03 (0.02)	0.03 (0.02)	0.293	0.02 (0.01)	0.02 (0.01)	0.720
ALB	42.80 (4.30)	41.70 (4.60)	0.005	41.80 (3.60)	42.40 (6.15)	0.731
GLU	4.75 (0.83)	4.87 (0.71)	0.295	4.72 (0.77)	4.71 (0.60)	0.518
UREA	5.00 (1.50)	5.30 (2.01)	0.65	5.50 (2.30)	5.60 (2.18)	0.655
CREA	62.00 (16.50)	53.00 (17.00)	<0.001	62.00 (17.00)	52.50 (32.00)	0.049
URIC	316.50 (84.00)	295.00 (104.00)	0.003	302.00 (100)	266.50 (92.75)	0.087
TG	1.33 (1.03)	1.13 (0.71)	0.007	1.29 (0.93)	1.08 (0.48)	0.004
CHOL	4.64 ± 0.85	4.61 ± 0.91	0.814	4.34 ± 0.73	4.88 ± 1.18	0.001
HDL_C	1.16 (0.42)	1.37 (0.46)	<0.001	1.19 (0.31)	1.37 (0.60)	<0.001
CK	180.00 (217.00)	145.00 (127.00)	0.002	142.00 (157.0)	111.50 (97.5)	0.074

SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; PLT, platelet count; NEUT#, neutrophil count; LYMPH#, lymphocyte count; PLR, the ratio of PLT to LYMPH#; MONO#, monocyte count; BASO#, basophil count; ALB, albumin; GLU, glucose; UREA, urea; CREA, creatinine; URIC, uric acid; TG, triglyceride; CHOL, cholesterol; HDL_C, high-density lipoprotein cholesterol.

the time of pulmonary function tests. High blood pressure was defined as systolic blood pressure (SBP) of ≥140 mmHg and/or diastolic blood pressure (DBP) of ≥90 mmHg in this study.

Hematologic and biochemical laboratory parameters were obtained using an automated hematology analyzer (KX-21 N, Sysmex America, Lincolnshire, IL, USA), and data on platelet count (PLT), neutrophil count (NEUT#), lymphocyte count (LYMPH#), monocyte count (MONO#), eosinophil count (EO#), basophil count (BASO#), albumin (ALB), glucose (GLU), triglyceride (TG), urea (UREA), cholesterol (CHOL), uric acid (URIC), high-density lipoprotein cholesterol (HDL_C), creatinine (CREA), and creatine kinase (CK) were obtained. PLR was defined as the ratio of PLT to LYMPH#. Using electronic spirometry, PEF was measured with the patient in a standing position.

Statistical Analysis

Correlation Analysis

Continuous variables with normal distribution were reported as mean (standard deviation [SD]); non-normal variables were presented as median (interquartile range (IQR), 25 to 75%). The linear correlations of these variables and the respiratory function index (PEF) were calculated by Spearman. For categorical variables, the correlations were performed by the nonparametric test (Wilcoxon rank-sum test). For normal distribution data, differences between the groups who had a PEF of <5.50 or

TABLE 2 | Variable in the prediction model by backward stepwise regression method combined with Akaike information criterion (AIC) for logistic regression equation.

variable	Estimate	Std.error	z-value	Degree of freedom	*P*-value
(Intercept)	−3.483	2.386	−1.46	1	0.144
Gender (female)	1.601	0.450	3.558	1	3.73E−04
Site.of.onset 2	1.373	0.505	2.718	1	0.007
Site.of.onset 3	0.695	1.239	0.561	1	0.575
BMI	−0.106	0.071	−1.499	1	0.134
PLT	0.005	0.003	1.857	1	0.063
'BASO#'	−14.968	10.708	−1.398	1	0.162
CREA	−0.041	0.014	−2.842	1	0.004
URIC	0.004	0.003	1.717	1	0.086
TG	0.366	0.234	1.562	1	0.118
CHOL	−0.483	0.236	−2.045	1	0.041
'HDL-C'	1.363	0.719	1.897	1	0.058
Onset.age	0.088	0.019	4.723	1	2.33E−06

Site.of.onset 2, bulbar onset; site.of.onset 3, the onset site except for limb onset and bulbar onset; BMI, body mass index; PLT, platelet count; BASO#, basophil count; CREA, creatinine; TG, triglyceride; URIC, uric acid; CHOL, cholesterol; HDL_C, high-density lipoprotein cholesterol.

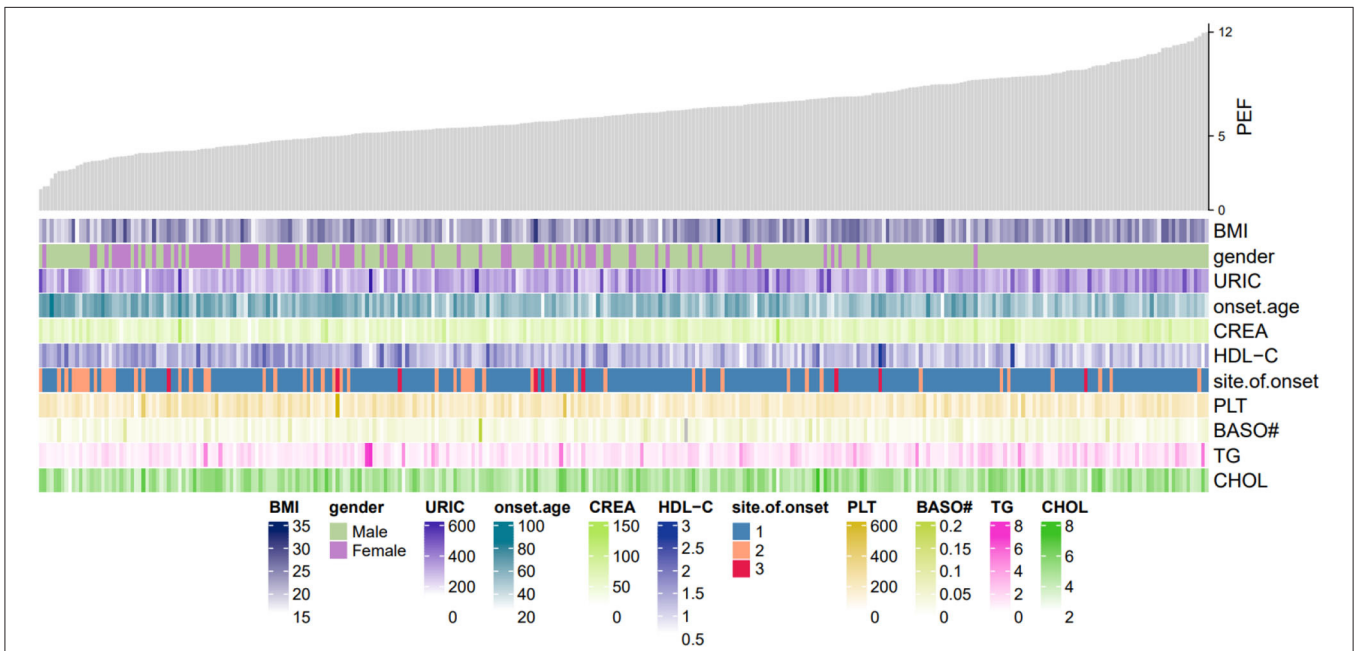


FIGURE 2 | Heatmap of PEF in this study. The gray shistogram represents values of PEF. Below the histogram are factors in the logistic regression model equation. The brightness of the color varied upon the value for continuous variables. PDF, peak expiratory flow; URIC, uric acid; BMI, body mass index; CREA, creatinine; HDL_C, high-density lipoprotein cholesterol; BASO#, basophil count; PLT, platelet count; TG, triglyceride; CHOL, cholesterol.

not were performed by the Independent Samples *t*-test. For non-normal distribution, differences were estimated using the Mann-Whitney test or chi-squared tests.

The Training Set and the Test Set

The retrospective cohort was divided into two independent sets: the training set and the test set. The training set included 70% (220/319) of the retrospective cohort, randomly selected from the cohort. The test set included 30% (95/319) of the retrospective cohort. Models were developed based on the characteristics of the training set. The test set was used to validate the models (**Figure 1**). Partition of the data set was realized by the sample() function in R. In the partitioning, the set.seed() function was used to generate a sequence of random numbers.

Logistic Regression

A total of 23 variables, namely, gender, onset age, site of onset, diagnostic level, height, BMI, Blood pressure, PLT, PLR, NEUT#, LYMPH#, MONO#, EO#, BASO#, ALB, GLU, UREA, CREA, URIC, TG, CHOL, HDL_C, and CK, were included to construct

the model. Based on the results of the backward stepwise regression method, combined the with Akaike information criterion (AIC), the heatmap of PEF for predicting peak expiratory flow impairment in patients with ALS was established. Receiver operating characteristic (ROC) curves and calibration curves were drawn to evaluate the predictive accuracy. Then, we tested the model in the validation set. All statistical analyses were performed using SPSS 24.0 (IBM, Chicago, IL, USA) and R 4.1.0 (www.rproject.org).

RESULTS

Of the 416 patients with Amyotrophic lateral sclerosis included in the study, there were 57 males (51.35%) in the retrospective cohort and 16 males (47.06%) in the prospective cohort with PEF of <5.5 L/s, while 54 females (48.65%) in the retrospective cohort and 18 females (52.94%) in the prospective cohort with PEF of <5.5L/s (**Table 1**). Similarly, there were significant age differences in onset age, BMI, level of PLR, CREA, TG, and HDL_C (all $P < 0.05$) when stratified by PEF either in the retrospective cohort or in the prospective cohort (**Table 1**).

We developed a multivariable logistic regression model for the association between the hematologic, biochemical laboratory parameters, clinical factors, and PEF in patients with ALS in the test set. The regression equation was developed from the estimate values obtained by z-value and is reported in **Table 2**. Moreover, we constructed a heatmap to show the relation between these variables in the equation and PEF (**Figure 2**).

Model PEF = −3.483+1.601 Gender (Female) + 1.373 Site.of.onset 2 (bulbar onset) or.695 Site.of.onset 3 (except for limb onset and bulbar onset) −0.106 BMI +0.005 PLT −14.968 'BASO#' −0.041 CREA level +0.004 URIC level +0.366 TG level −0.483 CHOL level + 1.363 HDL_C level +0.088 Onset.age (**Table 2**).

The ROC curve, with 11 predictive factors, revealed that the area under the curve was 80.2% in the test set (**Table 3**; **Figure 3**). And the ROC curve has a standard error of.049 with a 95% confidence interval (CI) of.707–0.898 in the test set. Furthermore, the calibration curve was also constructed (**Figure 4**). The calibration curves further indicated good calibration power.

TABLE 3 | ROC curve for the test set and the validation set.

	The test set	The validation set
The area under the ROC curve (AUC)	0.802	0.775
Standard error	0.049	0.052
95% Confidence interval	0.707–0.898	0.674–0.877
Z-statistic	6.197	5.326
Significance level P (Area=0.5)	5.77E−10	1.01E−07

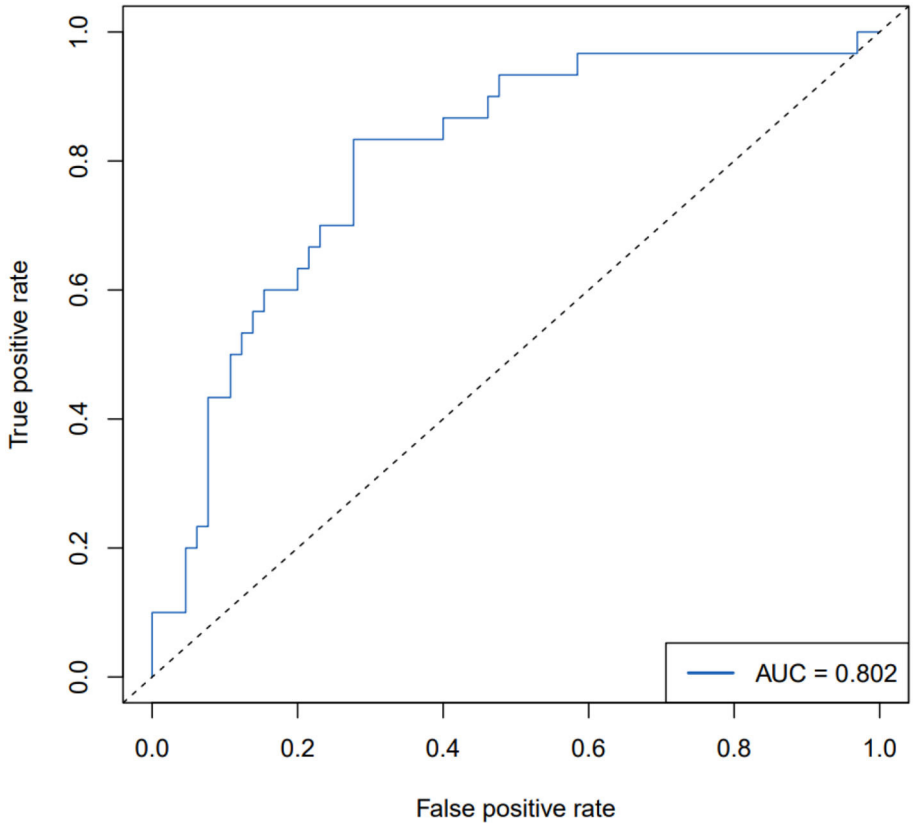


FIGURE 3 | ROC curve for the backward stepwise regression method combined with the AIC model for predicting PEF impairment in patients with ALS in the test set.

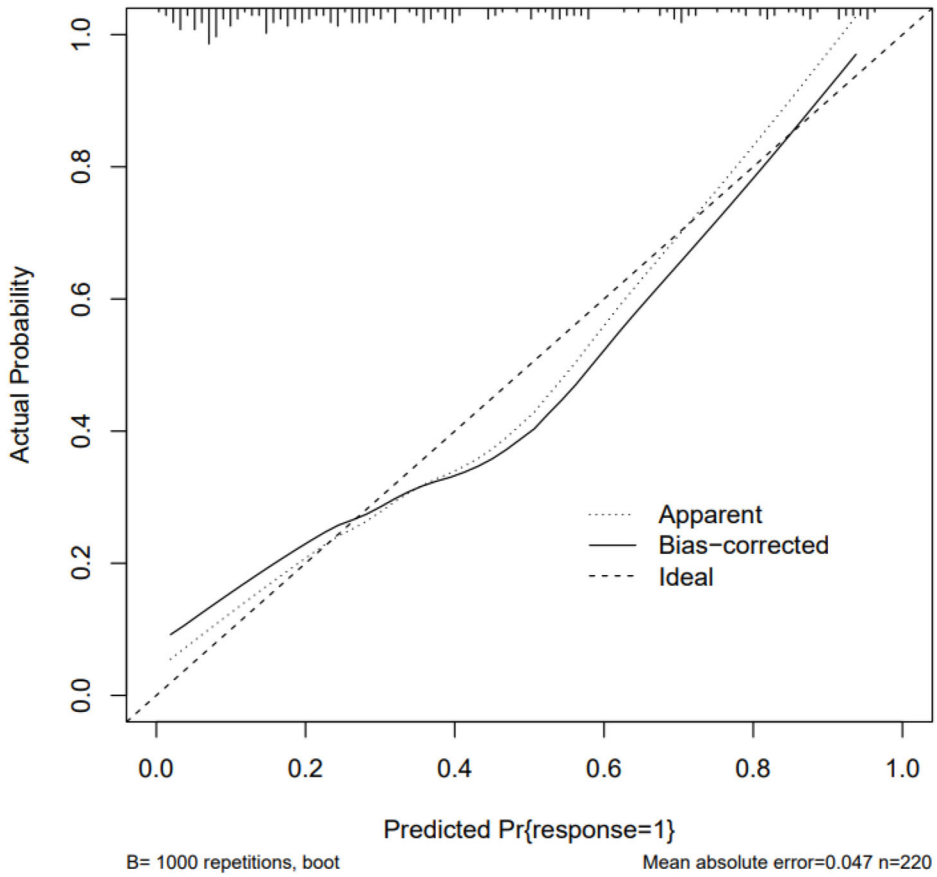


FIGURE 4 | The calibration curve for the model in the training set.

To validate the model, we tried to predict the data from the prospective cohort and calculated the ROC curve, which yielded a concordance statistic of.775 (95% CI,0.674–0.877) (**Figure 5**).

DISCUSSION

Since most patients with ALS in their advanced phase would develop respiratory impairment, it is important for clinicians to evaluate the respiratory function status. To date, there are few studies about models predicting respiratory function in patients with ALS (16, 23). Our study created and validated a multivariable logistic model for the association between PEF and the venous blood biomarkers and clinical factors in patients with ALS. Our single-site study revealed that PEF is related to ALS characteristics, as well as hematologic and biochemical laboratory parameters, and further verification is needed in external data sets with a large sample. To the best of our knowledge, this is the first study of a prediction model for PEF in ALS.

The PEF decreases in neuromuscular diseases (24) and showed excellent discriminant ability to evaluate the presence of aspiration pneumonia (13). Furthermore, the peak cough expiratory significantly correlates with bulbar impairment in patients with ALS (25) and maybe a good indicator to discriminate between ALS with "non-invasive ventilation (NIV) indication" and those with "no NIV indication yet" (26). The prediction model in our study consists of variables that are easy to get. Therefore, it can be a preliminary screening tool for the severity of the disease quickly and be used to evaluate the cough ability of patients with ALS in the outpatient department. The PEF prediction model may be a good screen tool for clinical utility.

In our study, patients with ALS with a lower level of TG, and a higher level of HDL_C showed a trend of reduced PEF. Also, in the prospective cohort, PEF was inversely associated with cholesterol. This study has not confirmed previous research, which found that lower serum lipid levels are related to reduced functional vital capacity (16). Another study from Japan reported that forced vital capacity (FVC) was inversely correlated with serum total cholesterol (17). Prospective cohort studies with a larger sample size are warranted to clarify the relationship between serum lipid and respiratory function.

In contradiction to previous research (23), we found that lower levels of creatinine and uric acid were associated with decreased PEF, but not independent predictors of PEF. Several

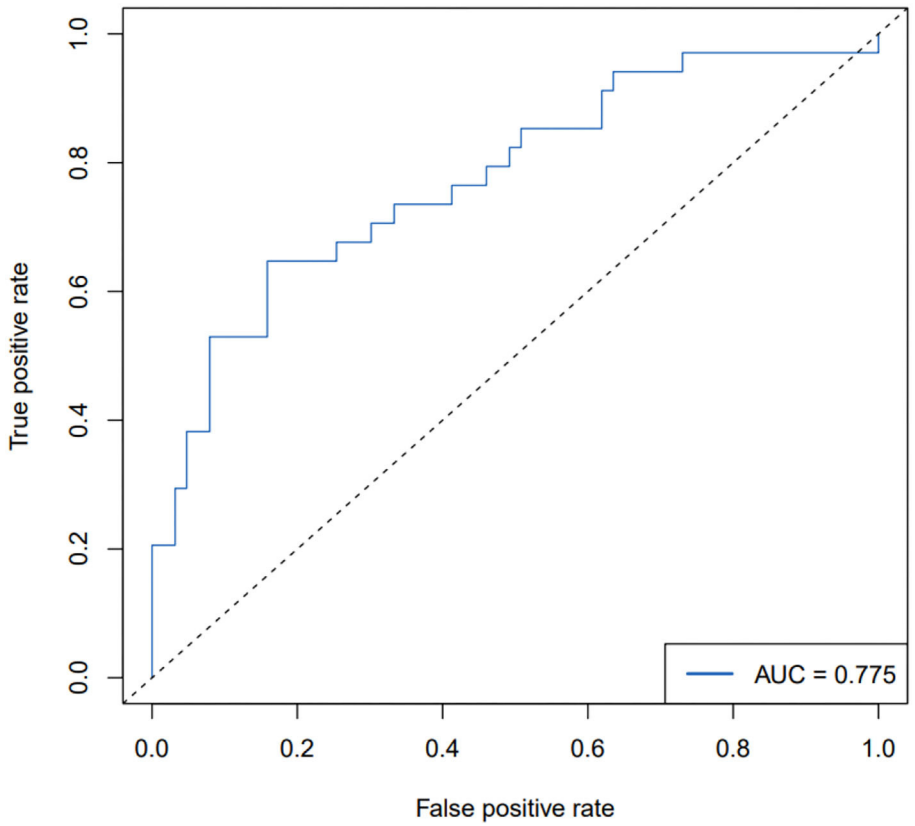


FIGURE 5 | Receiver operating characteristic (ROC) curve of PEF in the validation set.

studies have shown that low levels of serum creatinine and uric acid have been associated with poor survival, and faster functional decline (18–20, 27). One possibility is that patients with ALS with short survival have a faster respiratory decline. These patients have decreased levels of creatinine and uric acid at the same time, ultimately resulting in a low level of creatinine and uric acid-related to impaired respiratory.

Dysphagia in patients with ALS had a close relationship with respiratory impairment (13). A prospective study demonstrated that not only bulbar-onset but also most limb-onset patients with ALS had problems with dysphagia (6). Moreover, about 20% of spinal-onset patients with ALS, who did not show dysphagia at diagnosis, had initial respiratory impairment (28), highlighting the importance of evaluating PEF at the beginning of enrollment. In this study, female patients with ALS with bulbar-onset, later age at onset, and lower BMI tended to have reduced PEF in our study. Age at onset, onset region, and weight loss were factors independently influencing survival in patients with ALS (28–30). Further studies are needed to investigate the relationship between bulbar-onset, age at onset, BMI, and PEF.

Our results should be interpreted with caution due to several limitations. Firstly, our retrospective cohort could not analyze biomarkers potentially significant to the risk of PEF. For example, lipid and uric acid values collected at admission might be affected by the other conditions that coexisted with ALS, which we did not analyze. Secondly, the validation data set in our study was not technically an external data set, thus, external validation is needed. Thirdly, the prediction model is only suitable for probable and definite patients with ALS. Additionally, since the model is derived from the retrospective cohort and a single medical center, it should be tested and verified on larger prospective cohorts to evaluate the validity and predictability.

In conclusion, our study constructed a multivariable logistic regression equation of PEF derived from venous blood biomarkers and clinical factors in patients with ALS. The finding needs to be verified in prospective and multi-site studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Sichuan University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the

individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XH: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing-original draft, and writing-review and editing. JF, XC, QZ, QS, and FX: data curation, formal analysis, and methodology. HH: data curation, formal analysis, methodology, and funding acquisition. YX: conceptualization, funding acquisition, project administration, supervision, and writing-review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Basic Conditions Platform Construction Project of the Sichuan Science and Technology Department (2019JDPT0015), the 1.3.5 Project for disciplines of excellence, West China Hospital, Sichuan University (ZYJC18003), and the Miaozi Project in Science and Technology Innovation Program of Sichuan Province (2020JDRC0057).

ACKNOWLEDGMENTS

We thank the Amyotrophic lateral sclerosis individuals for their participation in our study.

REFERENCES

1. Kiernan M, Vucic S, Talbot K, McDermott C, Hardiman O, Shefner J, et al. Improving clinical trial outcomes in amyotrophic lateral sclerosis. *Nat Rev Neurol.* (2021) 17:104–18. doi: 10.1038/s41582-020-00434-z
2. Longinetti E, Fang F. Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Curr Opin Neurol.* (2019) 32:771–6. doi: 10.1097/WCO.0000000000000730
3. Hardiman O, Al-Chalabi A, Chio A, Corr E, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. *Nat Rev Dis primers.* (2017) 3:17071. doi: 10.1038/nrdp.2017.71
4. Andersen P, Abrahams S, Borasio G, de Carvalho M, Chio A, Van Damme P, et al. EFNS guidelines on the clinical management of amyotrophic lateral sclerosis (MALS)–revised report of an EFNS task force. *Eur J Neurol.* (2012) 19:360–75. doi: 10.1111/j.1468-1331.2011.03501.x
5. Halpin D, Meltzer E, Pisternick-Ruf W, Moroni-Zentgraf P, Engel M, Zaremba-Pechmann L, et al. Peak expiratory flow as an endpoint for clinical trials in asthma: a comparison with FEV. *Respir Res.* (2019) 20:159. doi: 10.1186/s12931-019-1119-6
6. Onesti E, Schettino I, Gori M, Frasca V, Ceccanti M, Cambieri C, et al. Dysphagia in amyotrophic lateral sclerosis: impact on patient behavior, diet adaptation, and riluzole management. *Front Neurol.* (2017) 8:94. doi: 10.3389/fneur.2017.00094
7. Tabor L, Gaziano J, Watts S, Robison R, Plowman E. Defining swallowing-related quality of life profiles in individuals with amyotrophic lateral sclerosis. *Dysphagia.* (2016) 31:376–82. doi: 10.1007/s00455-015-9686-2
8. Liu G, Su J, Wang X, He J. Value of peak expiratory flow rate in evaluating cough ability in patients undergoing lung surgery. *Can Respir J.* (2021) 2021:5888783. doi: 10.1155/2021/5888783
9. Wu Y, Zhou Y, Gao S, Du C, Yao L, Yang R. Effects of preoperative pulmonary function on short-term outcomes and overall survival after video-assisted thoracic surgery lobectomy. *Ann Transl Med.* (2021) 9:1651. doi: 10.21037/atm-21-5244
10. Chang S, Zhou K, Wang Y, Lai Y, Che G. Prognostic value of preoperative peak expiratory flow to predict postoperative pulmonary complications in surgical lung cancer patients. *Front Oncol.* (2021) 11:782774. doi: 10.3389/fonc.2021.782774
11. Borders J, Troche M. voluntary cough effectiveness and airway clearance in neurodegenerative disease. *J Speech Lang Hear Res.* (2021).65:431–49 doi: 10.1044/2021_JSLHR-21-00308
12. Tabor-Gray L, Vasilopoulos T, Wheeler-Hegland K, Wymer J, Plowman E. Reflexive airway sensorimotor responses in individuals with amyotrophic lateral sclerosis. *Dysphagia.* (2021) 36:574–82. doi: 10.1007/s00455-020-10171-6
13. plowman e, watts s, robison r, tabor l, dion c, gaziano j, et al. voluntary cough airflow differentiates safe vs. unsafe swallowing in amyotrophic lateral sclerosis. *Dysphagia.* (2016) 31:383–90. doi: 10.1007/s00455-015-9687-1
14. Pirola A, De Mattia E, Lizio A, Sannicolò G, Carraro E, Rao F, et al. The prognostic value of spirometric tests in Amyotrophic Lateral Sclerosis patients. *Clin Neurol Neurosurg.* (2019) 184:105456. doi: 10.1016/j.clineuro.2019.105456
15. Zhang Q, Huang J, Chen J, Hu W, Xu L, Guo Q. Peak expiratory flow is a reliably household pulmonary function parameter correlates with disease severity and survival of patients with amyotrophic lateral sclerosis. *BMC Neurol.* (2022) 22:105. doi: 10.1186/s12883-022-02635-z
16. Chiò A, Calvo A, Ilardi A, Cavallo E, Moglia C, Mutani R, et al. Lower serum lipid levels are related to respiratory impairment in patients with ALS. *Neurology.* (2009) 73:1681–5. doi: 10.1212/WNL.0b013e3181c1df1e
17. Ikeda K, Hirayama T, Takazawa T, Kawabe K, Iwasaki Y. Relationships between disease progression and serum levels of lipid, urate, creatinine and ferritin in Japanese patients with amyotrophic lateral sclerosis: a cross-sectional study. *Int Med.* (2012) 51:1501–8. doi: 10.2169/internalmedicine.51.7465
18. Chiò A, Calvo A, Bovio G, Canosa A, Bertuzzo D, Galmozzi F, et al. Amyotrophic lateral sclerosis outcome measures and the role of albumin and creatinine: a population-based study. *JAMA Neurol.* (2014) 71:1134–42. doi: 10.1001/jamaneurol.2014.1129
19. van Eijk R, Eijkemans M, Ferguson T, Nikolakopoulos S, Veldink J, van den Berg L. Monitoring disease progression with plasma creatinine in amyotrophic lateral sclerosis clinical trials. *J Neurol Neurosurg Psychiatry.* (2018) 89:156–61. doi: 10.1136/jnnp-2017-317077
20. Küffner R, Zach N, Norel R, Hawe J, Schoenfeld D, Wang L, et al. Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression. *Nat Biotechnol.* (2015) 33:51–7. doi: 10.1038/nbt.3051
21. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord.* (2000) 1:293–9. doi: 10.1080/146608200300079536
22. Agosta F, Al-Chalabi A, Filippi M, Hardiman O, Kaji R, Meininger V, et al. The El Escorial criteria: strengths and weaknesses. *Amyotroph Lateral Scler Frontotemporal Degener.* (2015) 16:1–7. doi: 10.3109/21678421.2014.964258
23. Morgadinho J, Pronto-Laborinho A, Conceição V, Gromicho M, Pinto S, Swash M, et al. Plasma creatinine level does not predict respiratory function in amyotrophic lateral sclerosis. *J Neuromuscul Dis.* (2021) 8:795–9. doi: 10.3233/JND-200583
24. Yamada S, Hashizume A, Hijikata Y, Inagaki T, Suzuki K, Kondo N, et al. Decreased peak expiratory flow associated with muscle fiber-type switching in spinal and bulbar muscular atrophy. *PLoS ONE.* (2016) 11:e0168846. doi: 10.1371/journal.pone.0168846
25. Enache I, Pistea C, Fleury M, Schaeffer M, Oswald-Mammosser M, Echaniz-Laguna A, et al. Ability of pulmonary function decline to predict death in amyotrophic lateral sclerosis patients. *Amyotroph Lateral Scler Frontotemporal Degener.* (2017) 18:511–8. doi: 10.1080/21678421.2017.1353097

26. Tilanus T, Groothuis J, TenBroek-Pastoor J, Feuth T, Heijdra Y, Slenders J, et al. The predictive value of respiratory function tests for non-invasive ventilation in amyotrophic lateral sclerosis. *Respir Res.* (2017) 18:144. doi: 10.1186/s12931-017-0624-8
27. Xu L, Hu W, Guo Q, Xu G, Wang N, Zhang Q. Serum Uric Acid Levels Predict Mortality Risk in Male Amyotrophic Lateral Sclerosis Patients. *Front Neurol.* (2021) 12:602663. doi: 10.3389/fneur.2021.602663
28. Moglia C, Calvo A, Grassano M, Canosa A, Manera U, D'Ovidio F, et al. Early weight loss in amyotrophic lateral sclerosis: outcome relevance and clinical correlates in a population-based cohort. *J Neurol Neurosurg Psychiatry.* (2019) 90:666–73. doi: 10.1136/jnnp-2018-319611
29. Mandrioli J, Malerba S, Beghi E, Fini N, Fasano A, Zucchi E, et al. Riluzole and other prognostic factors in ALS: a population-based registry study in Italy. *J Neurol.* (2018) 265:817–27. doi: 10.1007/s00415-018-8778-y
30. Pinto S, de Carvalho M. Comparison of slow and forced vital capacities on ability to predict survival in ALS. *Amyotroph Lateral Scler Frontotemporal Degener.* (2017) 18:528–33. doi: 10.1080/21678421.2017.1354995

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 He, Feng, Cong, Huang, Zhao, Shen, Xu and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

SYSTEMATIC REVIEW
published: 27 May 2022
doi: 10.3389/fpubh.2022.884645

A Retrospective Study of Climate Change Affecting Dengue: Evidences, Challenges and Future Directions

Surbhi Bhatia[1], Dhruvisha Bansal[2], Seema Patil[2], Sharnil Pandya[2], Qazi Mudassar Ilyas[1] and Sajida Imran[3]*

[1] Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia, [2] Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, India, [3] Department of Computer Engineering, College of Computer Sciences and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia

OPEN ACCESS

Edited by:
Celestine Iwendi, School of Creative Technologies University of Bolton, United Kingdom

Reviewed by:
Suyel Namasudra, Universidad Internacional de La Rioja, Spain
Jude Osamor, Glasgow Caledonian University, United Kingdom
Praise Young, Nnamdi Azikiwe University, Nigeria

***Correspondence:**
Surbhi Bhatia
sbhatia@kfu.edu.sa

Specialty section:
This article was submitted to Frontiers in Public Health, a section of the journal Frontiers in Public Health

Received: *26 February 2022*
Accepted: *26 April 2022*
Published: *27 May 2022*

Citation:
Bhatia S, Bansal D, Patil S, Pandya S, Ilyas QM and Imran S (2022) A Retrospective Study of Climate Change Affecting Dengue: Evidences, Challenges and Future Directions. Front. Public Health 10:884645. doi: 10.3389/fpubh.2022.884645

Climate change is unexpected weather patterns that can create an alarming situation. Due to climate change, various sectors are affected, and one of the sectors is healthcare. As a result of climate change, the geographic range of several vector-borne human infectious diseases will expand. Currently, dengue is taking its toll, and climate change is one of the key reasons contributing to the intensification of dengue disease transmission. The most important climatic factors linked to dengue transmission are temperature, rainfall, and relative humidity. The present study carries out a systematic literature review on the surveillance system to predict dengue outbreaks based on Machine Learning modeling techniques. The systematic literature review discusses the methodology and objectives, the number of studies carried out in different regions and periods, the association between climatic factors and the increase in positive dengue cases. This study also includes a detailed investigation of meteorological data, the dengue positive patient data, and the pre-processing techniques used for data cleaning. Furthermore, correlation techniques in several studies to determine the relationship between dengue incidence and meteorological parameters and machine learning models for predictive analysis are discussed. In the future direction for creating a dengue surveillance system, several research challenges and limitations of current work are discussed.

Keywords: predictive models, machine learning, surveillance system, dengue, climatic factors

1. INTRODUCTION

Artificial Intelligence is a field in which a machine exhibits intelligence by learning about itself. Artificial Intelligence may employ various approaches and algorithms to comprehend human intellect, but it is not limited. Machine Learning is a branch of AI that deals with techniques that can learn from experience on their own (1). In automated decision-making and predictive analytics, AI plays a crucial role (2). Methods based on various kinds of AI are already being used in a variety of climate change and environmental monitoring research sectors (3). Machine Learning and Deep Learning

are used in education, finance, healthcare, agriculture, and many more. The application of Machine Learning in healthcare demonstrated encouraging results, instilling trust in the sector (4). There are varied reasons for using artificial intelligence in healthcare. Primarily, it helps to manage the clinical records of patients and aids in providing personalized treatment and medicines, early identification of disease, providing predictive analytics, which could help health officials take preventive measures in advance before the outbreak of any disease (5). Artificial Intelligence goes back to the too early 1990s and today has evolved into a burgeoning field with a significant contribution in healthcare (6) through various modes like drug and discovery, personalized treatment and medicine, preventive measures, a health assistant chatbot, clinical trial and researches, diagnosing and pathology, analytics, image processing and prediction of an outbreak. Artificial Intelligence is a boon, and **Figure 1** demonstrates the evolution of AI in healthcare throughout 1950–2020. Climate change is long-term changes in temperature and weather patterns.These variations might be attributed to natural causes such as solar cycle oscillations. Due to the use of fossil fuels such as coal, oil, and gas, human activities have been the leading cause of climate change since the 1800s (7). Greenhouse gas emissions from fossil fuel combustion act as a blanket, trapping the sun's heat and boosting global temperatures. The amount of greenhouse gases in the atmosphere is at its highest level in 2 million years. Artificial Intelligence is a field in which a machine exhibits intelligence by learning about itself. Artificial Intelligence may employ various approaches and algorithms to comprehend human intellect, but it is not limited. Machine Learning is a branch of AI that deals with techniques that can learn from experience on their own. In automated decision-making and predictive analytics, AI plays a crucial role. Methods

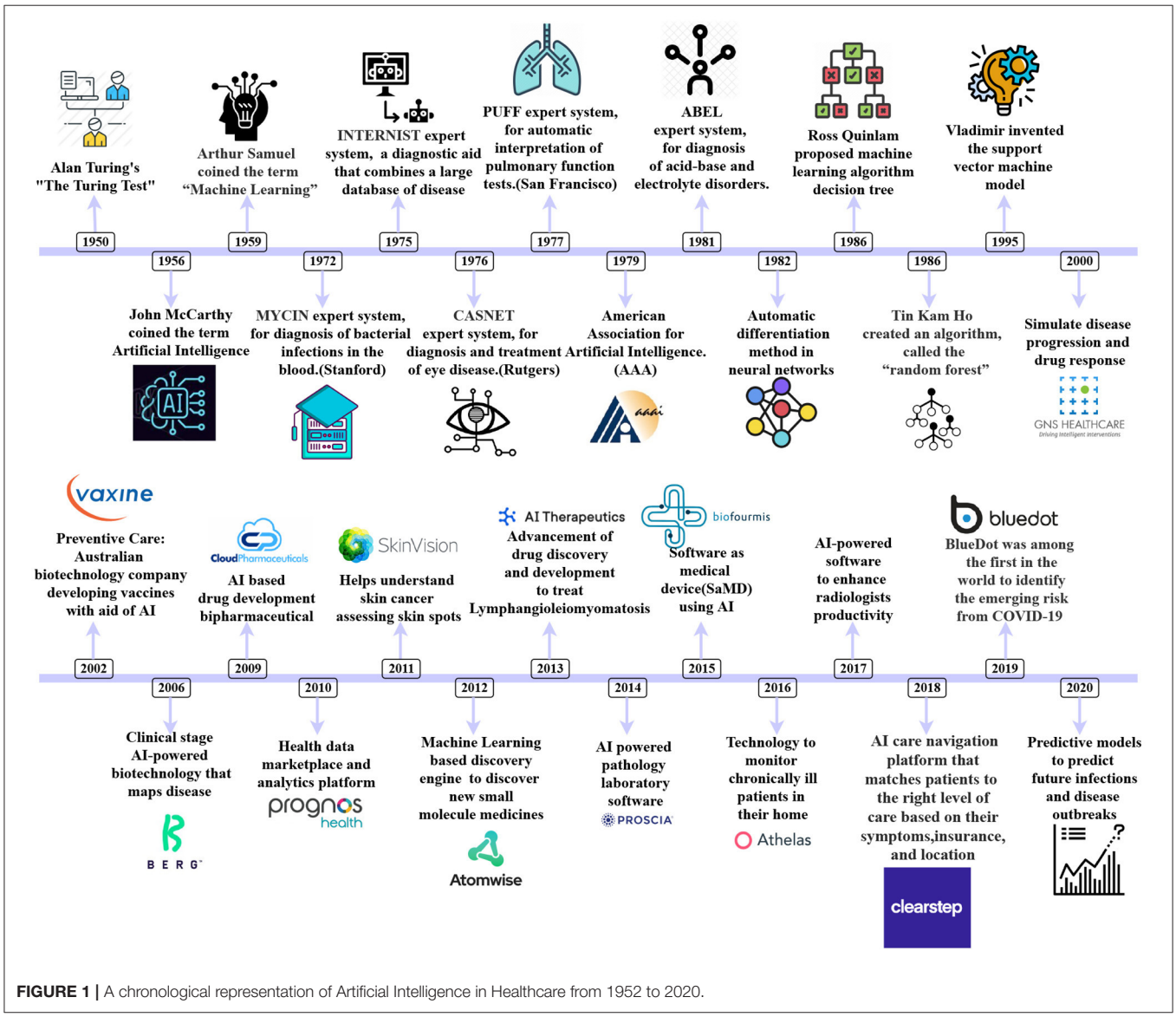


FIGURE 1 | A chronological representation of Artificial Intelligence in Healthcare from 1952 to 2020.

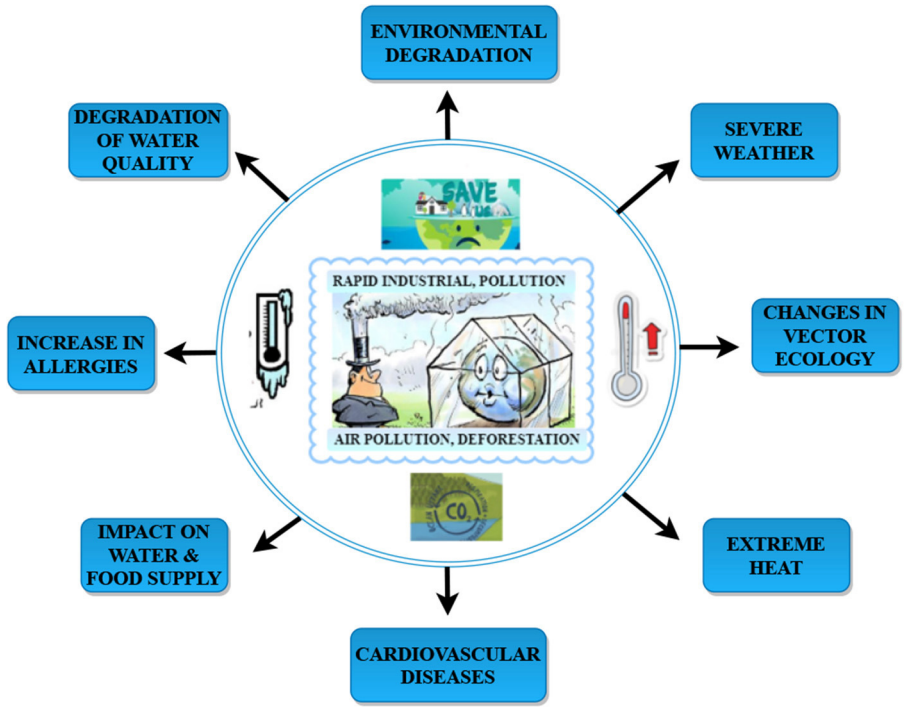


FIGURE 2 | Causes and consequences of climate change.

based on various kinds of AI are already being used in a variety of climate change and environmental monitoring research sectors. Machine Learning and Deep Learning are used in education, finance, healthcare, agriculture, and many more. The application of Machine Learning in healthcare demonstrated encouraging results, instilling trust in the sector (8). There are varied reasons for using artificial intelligence in healthcare. Primarily, it helps to manage the clinical records of patients and aids in providing personalized treatment and medicines, early identification of disease, providing predictive analytics, which could help health officials take preventive measures in advance before the outbreak of any disease. Artificial Intelligence goes back to the too early 1990s and today has evolved into a burgeoning field with a significant contribution in healthcare through various modes like drug and discovery, personalized treatment and medicine, preventive measures, a health assistant chatbot, clinical trial and researches, diagnosing and pathology, analytics, image processing and prediction of an outbreak. Artificial Intelligence is a boon, and **Figure 1** demonstrates the evolution of AI in healthcare throughout 1950–2020. Climate change is long-term changes in temperature and weather patterns.These variations might be attributed to natural causes such as solar cycle oscillations. Due to the use of fossil fuels such as coal, oil, and gas, human activities have been the leading cause of climate change since the 1800s. Greenhouse gas emissions from fossil fuel combustion act as a blanket, trapping the sun's heat and boosting global temperatures. The amount of greenhouse gases in the atmosphere is at its highest level in 2 million years. Emissions are steadily increasing. Since the late 1800s, the Earth has warmed by around 1.1°C. The previous 10 years (2011–2020) have been the hottest on record. **Figure 2** explains the causes of climate change and the effects due to those variations. Vector-borne illness is a severe public health threat in underdeveloped nations that is only becoming worse. Temperatures in the air and water, precipitation patterns, severe rainfall events, and seasonal changes are all known to impact disease transmission (9). At the moment, dengue (10) is spreading widely after COVID-19; dengue is a vector-borne disease that spreads through an infected female mosquito of species Aedes aegypti (11) and is one of the most common diseases in more than 100 countries (12). Aedes aegypti can be found in urban and suburban settings with high human population density and housing density. This species is endophilic, which means it seeks shelter inside structures. As a result, the intradomicile is more typically seen than the peridomicile. Tires, cans, bottles, pots, vats, brass, swimming pools, and abandoned aquariums are frequent breeding habitats filled with rainwater or household water.

These mosquitos carry the Chikungunya, yellow fever, and Zika viruses. Dengue fever is widespread in the tropics, with different risk levels based on rainfall, temperature, relative humidity, and unplanned fast urbanization. Dengue fever may be fatal if it is not treated appropriately. It was first discovered during dengue epidemics in the Philippines and Thailand in

the 1950s. Most Asian and Latin American nations now have severe dengue fever, a leading cause of hospitalization and mortality among children and adults. A Flaviviridae virus causes dengue with four unique but closely related serotypes (DENV-1, DENV-2, DENV-3, DENV-4), and the fifth serotype of dengue has been detected in the Malaysian state of Sarawak (13). Dengue fever is caused by DENV infection in our bodies. **Figure 3** depicts the common symptoms of dengue. Dengue fever causes severe joint and muscle pain, enlarged lymph nodes, headaches, fever, exhaustion, and a rash. Dengue fever is characterized by non-specific flu-like symptoms such as chills, appetite loss, lethargy, and low backache. According to a recent WHO research, dengue fever grew in several countries in 2020, including Bangladesh, Brazil, the Cook Islands, Ecuador, India, Indonesia, the Maldives, Mauritania, Mayotte, Nepal, Singapore, Sri Lanka, Sudan, Thailand, Timor-Leste, and Yemen. In 2019,

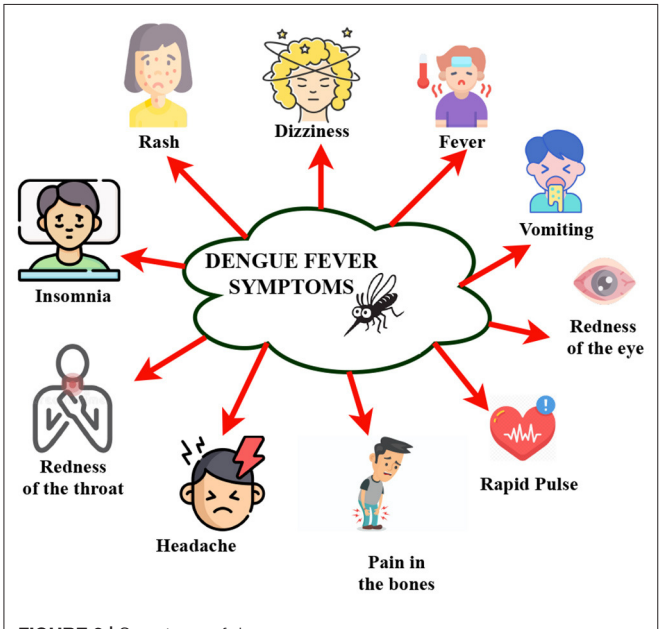


FIGURE 3 | Symptoms of dengue.

dengue fever cases reached an all-time high in the world. Dengue fever is slowly gaining ground in 2021, with instances documented in several nations. Around 40 per cent of the world's population lives in areas with a high risk of dengue fever, such as tropical and subtropical climates. Dengue fever has become much more common in recent decades worldwide. According to one estimate, 390 million dengue virus infections occur each year (12). Emissions are steadily increasing. Since the late 1800s, the Earth has warmed by around 1.1°C. The previous 10 years (2011–2020) have been the hottest on record (9). **Figure 2** explains the causes of climate change and the effects due to those variations. Vector-borne illness is a severe public health threat in underdeveloped nations that is only becoming worse. Temperatures in the air and water, precipitation patterns, severe rainfall events, and seasonal changes are all known to impact disease transmission (10). At the moment, dengue (11) is spreading widely after COVID-19; dengue is a vector-borne disease that spreads through an infected female mosquito of species Aedes aegypti (12) and is one of the most common diseases in more than 100 countries (14). Aedes aegypti can be found in urban and suburban settings with high human population density and housing density. This species is endophilic, which means it seeks shelter inside structures. As a result, the intradomicile is more typically seen than the peridomicile. Tires, cans, bottles, pots, vats, brass, swimming pools, and abandoned aquariums are frequent breeding habitats filled with rainwater or household water (13). These mosquitos carry the Chikungunya, yellow fever, and Zika viruses. Dengue fever is widespread in the tropics, with different risk levels based on rainfall, temperature, relative humidity, and unplanned fast urbanization. Dengue fever may be fatal if it is not treated appropriately. It was first discovered during dengue epidemics in the Philippines and Thailand in the 1950s. Most Asian and Latin American nations now have severe dengue fever, a leading cause of hospitalization and mortality among children and adults. A Flaviviridae virus causes dengue with four unique but closely related serotypes (DENV-1, DENV-2, DENV-3, DENV-4), and the fifth serotype of dengue has been detected in the Malaysian state of Sarawak (15). Dengue fever is caused by

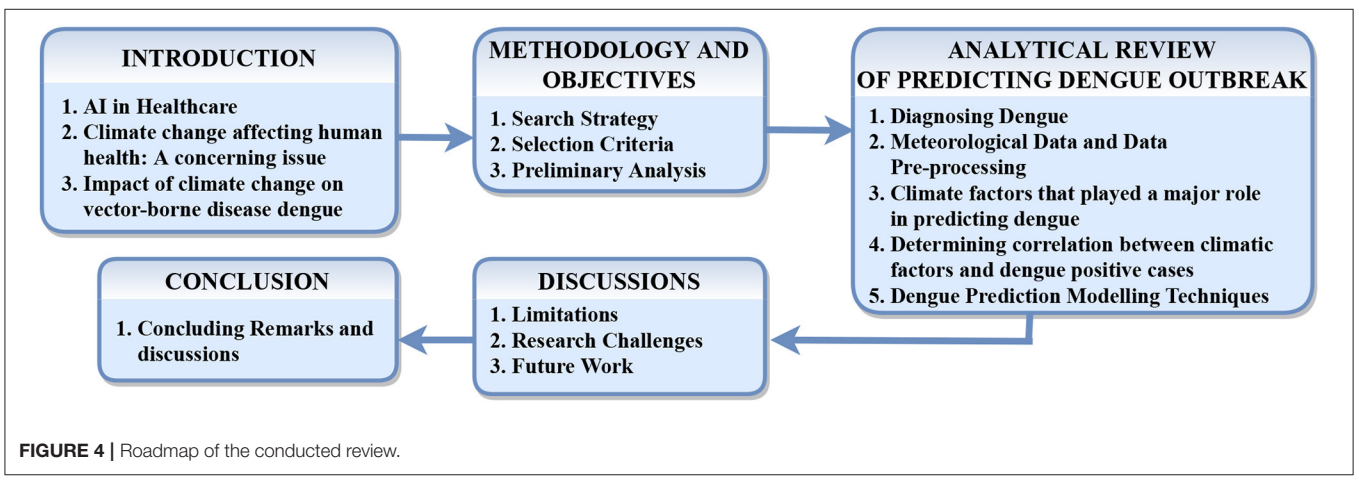


FIGURE 4 | Roadmap of the conducted review.

DENV infection in our bodies. Dengue fever causes severe joint and muscle pain, enlarged lymph nodes, headaches, fever, exhaustion, and a rash. Dengue fever is characterized by non-specific flu-like symptoms such as chills, appetite loss, lethargy, and low backache. According to a recent WHO research, dengue fever grew in several countries in 2020, including Bangladesh, Brazil (16), the Cook Islands, Ecuador, India (17), Indonesia (18), the Maldives, Mauritania, Mayotte, Nepal, Singapore (19), Sri Lanka, Sudan, Thailand (20), Timor-Leste, and Yemen. In 2019, dengue fever cases reached an all-time high in the world. Dengue fever is slowly gaining ground in 2021, with instances documented in several nations (14). Around 40 per cent of the world's population lives in areas with a high risk of dengue fever, such as tropical and sub-tropical climates. Dengue fever has become much more common in recent decades worldwide. According to one estimate, 390 million dengue virus infections occur each year (14). Dengue fever will affect 60 per cent of the world's population by 2080, according to researchers who blame climate change for the disease's spread. Climate change is widely cited as a contributing factor in the rapid spread of pandemic disease (21). Climate change is frequently identified by WHO, national, and international health authorities as one of the primary causes of the global spread of dengue fever and other Aedes-transmitted viral infections. According to the World Health Organization, dengue fever has become increasingly widespread internationally in recent decades (WHO). According to the World Health Organization, dengue fever affects 50–100 million people globally each year. As a result, it is vital to anticipate dengue epidemics. The accuracy of dengue epidemic prediction is currently a problem that must be addressed. The role of climate variables in predicting dengue outbreaks has been studied in a small number of studies. Tropical countries are the hardest hit because of their environmental, climatic, and socioeconomic characteristics (22). The weather has an impact on the vector-borne illness dengue's temporal and geographical spread. As a result, rainfall and ambient temperature are referred to as macro factors that influence dengue since they directly impact Aedes aegypti population density, which varies seasonally based on these important variables. Its population density tends to decline drastically during less precipitation and lower average temperatures in regions with a tropical or subtropical climate. Still, it grows consistently in locations with a tropical or subtropical environment (23). As the vector-weather association is as significant as the vector-human interaction, studies of climatic variables can have a better understanding of epidemic seasonality and the ability to anticipate it (24). In recent years, epidemiological research has focused on developing mathematical and statistical models based on weather parameters to explain the dynamics of dengue fever incidence. Its main goal was to find models with promising future predicting the potential of dengue incidence to help public health officials. Several researchers investigate the link between climate variables and dengue fever, frequently employing time-series analyses to characterize temporal trends, uncover patterns, and even make forecasts. Bhatt et al. (21) Many parameters, such as entomological, epidemiological, and geographic characteristics, can help predict the dengue outbreak effectively. Artificial intelligence has been employed in healthcare for a long time. This review study discusses the studies undertaken until now. **Figure 4** represents road-map of the present study. It investigates several other vulnerable groups that will be useful in predicting positive dengue cases and eventually help prevent the outbreak and take preventive measures at an early stage. The present study is structured as follows: Section 2 describes the methodology and objectives of the survey carried out; Section 3 provides a detailed analytical review of the papers; Section 4 discusses the challenges and future work, and Section 5 concludes the present study.

2. METHODOLOGY AND OBJECTIVES

To carry out the systematic literature review of dengue prediction models, process was followed which included i) establishing research questions, ii) developing a search strategy to narrow down the results, iii) selecting papers using eligibility criteria and iv) analyzing to extract strengths, weaknesses, and difficulties to overcome from articles. The main objectives of this systematic literature review are i) To gather and describe dengue machine learning models for dengue surveillance systems and ii) To illustrate the obstacles that future dengue modeling studies will face.

2.1. Search Strategy

Various digital libraries like Taylor and Francis, Google Scholar, MDPI, IEEE Xplorer, ScienceDirect, Pubmed, BMC and PLOS were used. The inclusion criteria were articles from January 2017 to August 2021 related to various dengue prediction models based on explanatory variables. The requirements for discarding the publications were personal opinions, unpublished works, posters, short articles and conference papers. The selection procedure included choosing themes by scrutinizing their title, keywords and abstract to eliminate the results that were not related. Further, to investigate more relevant papers, we examined summaries to determine if we could select the article or not.

2.2. Preliminary Analysis

To analyze the amount of research carried out, we analyzed Scopus data using keywords dengue, predicting, climatic factors, age and gender. **Figure 5A** were graphically represented using the query "[(predicting dengue) and (climatic factors and dengue)]" which fetched 1,315 documents, and **Figure 5B** were graphically represented using the query [(predicting dengue) AND (climatic factors and dengue) and (age group and gender)] which fetched only 38 documents. Therefore, we can deduce that studies on dengue prediction based on climatic factors, age group and gender has not been done by many researchers, and we can explore this field. As it is not carried out in many regions, there is a scope to examine countries/territories where this study has not been carried out. The analytic review that we carried out had the highest number of studies were from India, Brazil and Malaysia. **Figure 6** displays the number of studies done in different countries and republics. As stated by the WHO report, these countries belong to tropical and sub-tropical regions that are more prone to

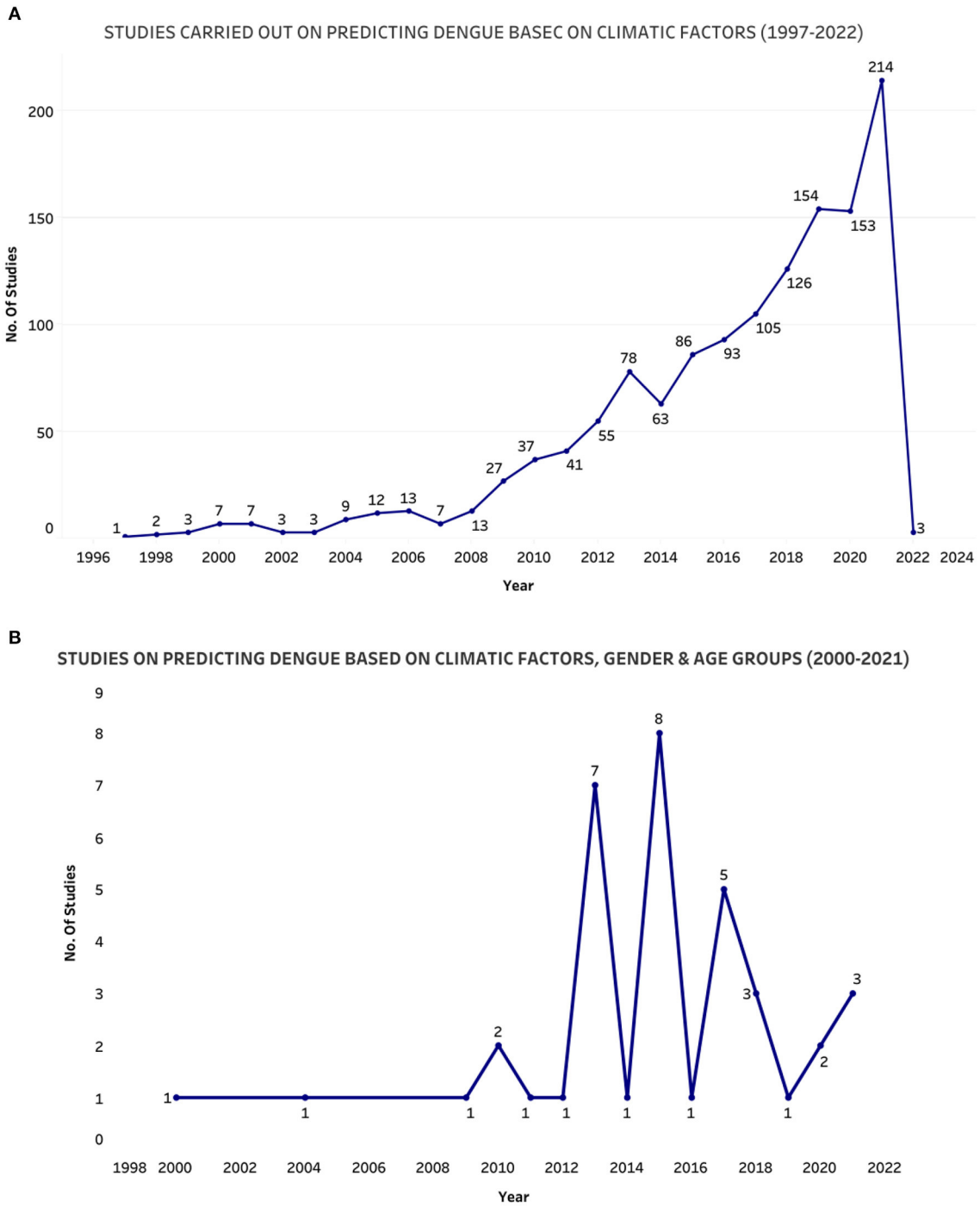


FIGURE 5 | No. of studies in different regions and time-period of scopus database: **(A)** No. of studies carried out on predicting dengue based on climatic factors for the time period 1997–2022 and **(B)** No. of studies carried out on predicting dengue based on climatic factors, age group and gender for the time-period 2000–2021.

dengue. Various explanatory variables are presently used for dengue modeling, categorized according to their characteristics and method of collection. The variables used were clinical, economic, laboratory and climatic factors. In general, climatic factors mainly were used. According to this review, vulnerable groups like age group and gender were least for dengue outbreak prediction as data about patients is not that easily available.

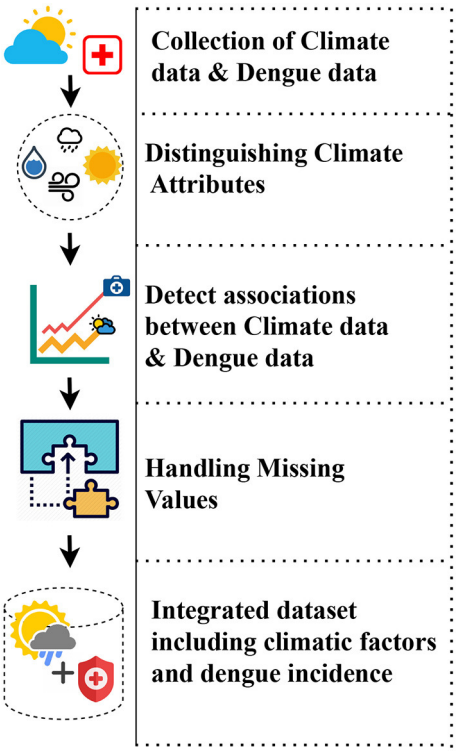


FIGURE 6 | Data pre-processing stages.

3. ANALYTICAL REVIEW OF PREDICTING DENGUE OUTBREAK

Forty articles of period 2019–2021 were reviewed and analyzed to understand the dengue prediction models and how different factors like meteorological conditions, laboratory tests, symptoms, demographic features were dependent variables for the rise of dengue cases.

3.1. Diagnosing Dengue

Identifying the dengue infection, viral antigens, viral RNA, and antibodies against the virus in the patient's blood or tissues are among the laboratory procedures used to diagnose dengue. Only 4–5 days after symptoms can the virus be discovered in the blood (25). Dengue can be diagnosed by isolating the virus, viral RNA, and viral protein during the early stages of the disease. Identifying antibodies IgM and IgG in an infected person's blood is an indirect approach to diagnosing dengue fever. This approach is widely used to identify dengue fever in its latter stages after virus levels have dropped. Doctors can discover antibodies to dengue 5 days after the onset of symptoms in most individuals, and IgG can be detected for months or even years after infection. IgM levels are very high after a primary (initial) dengue infection, but they are decreased with subsequent infection. During a subsequent infection, IgG levels increase. As a result, doctors can use IgM and IgG concentrations to determine whether a patient has a primary or secondary dengue infection (26). Therefore, in most of the data, IgG and IgM and levels are specified for easy detection and also it can aid in creating the laboratory dengue surveillance system (27).

3.2. Meteorological Data and Data Pre-processing

Meteorological data represents the weather conditions like relative humidity, precipitation, minimum temperature, maximum temperature, air pressure, wind speed and many other parameters for location and time. These are massive data collected and later utilized to record the highest and lowest climatic events for various reasons like weather forecasting to identify seasonal change and public health. There are various sources through which we can obtain data. A brief description of the explanatory variables, data source, region and period is presented in the **Table 1** used in different studies. Data Pre-processing is a crucial stage after data collection. This step enhances the model's performance, to which they will further provide the data for training and testing. Dengue incidence data collected contains information that might not be required, and there might be missing values. To handle these shortcomings, data pre-processing is a significant step. **Figure 6** shows the steps followed in information pre-processing. The major concern for reviewers is missing data and data not being available completely (28). Jorge et al. (29) developed a three-stage to take the missing values in the dataset wherein first stage parameters with more than 20 per cent of missing values were eliminated. At the next stage, they discarded cases with a non-response rate higher than 80 per cent. There are no particular rules for selecting the proper imputation of missing variables. It relies on the dataset type, non-response type, pattern of loss of response, research aims, specific population features, general study organization characteristics, or available software. Given the data entry procedure's features and the dataset's epidemiological nature, they concluded that imputing the mean of the valid neighboring data to the missing values was the best option at the third stage (30).

3.3. Climate Factors That Played a Major Role in Predicting Dengue

According to many studies, certain meteorological factors like temperature, rainfall, humidity, wind speed, air pressure and vegetation index displayed a positive correlation with the rise in dengue incidence. Hence, it helps us narrow down to a few concrete factors that can be used in further research to create dengue surveillance systems for different vulnerable groups. **Figure 7** displays the percentage and combination of climatic factors that have been used in review papers, where temperature, rainfall and humidity account for 49 per cent. Therefore, other least used elements that show some relation with dengue incidence can be used for efficient prediction.

3.3.1. Temperature

The most critical weather component affecting mosquito vector growth and dispersion and a potential predictor of dengue outbreak is temperature (31). The density of the mosquito population, biting rates, gonotrophic cycle lengths, and vector

TABLE 1 | Detailed representation of explanatory variables and data sources of different studies.

References	Region	Period	Explanatory variables/data source
Leandro et al. (15)	Brazil	2007–16	Dengue Data: State Department of Health of the State of Rio de Janeiro Temperature: Moderate Resolution Imaging Spectroradiometer Rainfall : Tropical Rainfall Measuring Mission
Joseph et al. (27)	Philippines	2014–18	Age and Gender of patient: Philippine Epidemiological Bureau (Department of Health)
Daniel et al. (39)	Mexico	2010–14	Dengue Incidence Count/Ministry of Health of the State Government of San Luis Potosí Environmental (Temperature, Rainfall, Relative Humidity and Elevation above sea level): INEGI (National Institute of Statistics and Geography - Mexico) Proximity (Distance to water bodies, vegetation and roads): UGSG Social (Population, Piped Water Index, Drainage Index, Human Development Index): INEGI
Jian Cheng et al. (49)	China	2006–15	Daily number of dengue cases: China Centre for Disease Control and Prevention Temperature, Relative Humidity and Rainfall: National Meteorological Information Center
HaorongMeng et al. (50)	China	2006–18	Dengue Case Count: Chinese National Notifiable Infectious Disease Reporting Information System Temperature and Precipitation: China Meteorological Data Service Center
Kumar Shashvat et al. (51)	India	2014–17	Dengue Cases: Integrated Disease Surveillance Programme, Government of India. Rainfall and Relative Humidity: Indiastat.com
Sourabh Bal et al. (38)	India	2005–16	Records of dengue cases/Directorate of Health Services, Government of West Bengal Temperature, Relative Humidity and Rainfall: IMD
Pi Guo et al. (52)	China	2011–14	Dengue case count: China National Notifiable Disease Surveillance System Temperature, Relative Humidity and Rainfall: China Meteorological Data Sharing Service System
Wei Wu et al. (53)	China	2003–14	Dengue Cases: National Notifiable Infectious Disease Reporting Information System (NIDRIS) Temperature, Relative Humidity and Rainfall: WorldClim Population Density, Road Density, Land use and cover: RESDC / GeoSOS
Sabrina Islaml et al. (28)	Bangladesh	2002–13	Dengue Case Data: Directorate General of Health and Services Temperature, Relative Humidity and Rainfall: Bangladesh Meteorological Department
Teerawad Sriklin et al. (32)	Thailand	2015–19	Dengue Fever Cases: Bureau of Epidemiology, Ministry of Public Health Temperature, Rainfall, Air Pressure and relative humidity: Meteorological Department of Southern Thailand
Gayan et al. (54)	Sri Lanka	2005–17	Dengue Incidence Data: Regional Epidemiology Unit Temperature, Relative Humidity and Rainfall: Department of Meteorology, Colombo
Felestin et al. (55)	Malaysia	2010–13	Dengue Fever Confirmed Cases: Ministry of Health Malaysia (MOH) portal Temperature, Rainfall and Relative Humidity: Malaysian Meteorological Department

size are all affected by temperature during the mosquito's incubation life cycle and behavior. The findings of Teerawad Sriklin et al. study.'s back with prior research suggest that severe temperatures alter the development of dengue vectors. As a result, the temperature has an impact on vectorial efficiency as well as the probability of an epidemic. Given the link between temperature and dengue cases, the predicted temperature change could worsen dengue transmission (28, 32).

3.3.2. Rainfall

Dengue transmission was found to be linked to rain and wet days. Rainfall has been identified as a role in the spread of dengue fever. Mosquitoes spend their whole life cycle in water before hatching into adult mosquitoes. Increased rainfall may create new habitat for larvae and vectors and boost adult survival (33).

3.3.3. Humidity

Despite widespread interest in the relationship among climatic conditions and positive dengue cases prevalence among researchers, research on relative humidity as a critical climatic component has been limited. Furthermore, the findings of the few study were equivocal and contradictory. They showed relative humidity as the most crucial predictor in an Indonesian investigation of dengue outbreaks in that nation, with a 3–4-month lag time. According to this study, low humidity in September and October is generally followed by a dengue

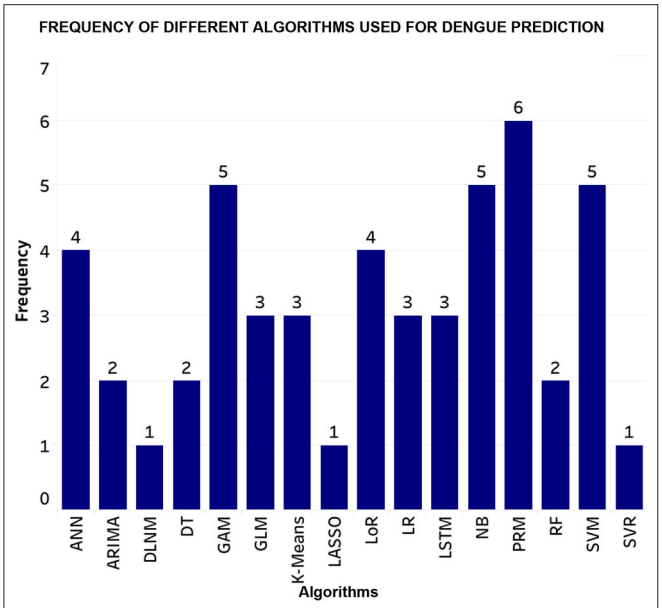


FIGURE 7 | Frequency of machine learning techniques used for dengue incidence prediction (SVM, Support Vector Machine; NB, Naive Bayes; DLNM, Distributed Lag Non-Linear Model; PRM, Poisson Regression Model; LSTM, Long Short Term Memory, ANN, Artificial Neural Network; GLM, Generalized Linear Model; GAM, Generalized Additive Model; LoR, Logistic Regression; LR, Linear Regression; RF, Random Forest; DT, Decision Tree; SVR, Support Vector Regressor).

outbreak early the following year. As a result, it's highly likely that if seasonal circumstances vary as a result of climate change, seasonal dengue outbreaks will shift as well (34).

3.3.4. Geographic Factors

The influence of residential regions on mosquito incidence became more evident under certain environmental circumstances, such as lower precipitation. According to research, mosquito incidence is very susceptible to high residential sites with a more significant density of residential streets. Roads feature drainage components that collect surface water runoff and release it in appropriate areas to avoid inland floods. However, concrete structures encroaching on channels or garbage filling them can often undermine adequate drainage in residential areas. These obstructions can prevent complete water flow, resulting in water accumulation places that can provide favorable habitat for Aedes aegypti. A growing body of research suggests a link between drainage and the presence of Aedes aegypti (28).

3.4. Determining Correlation Between Climatic Factors and Dengue Positive Cases

Correlation is a statistical method for deciding the degree to which two parameters are related, using correlation coefficients. Correlation coefficients are used to determine the strength of a linear relationship between two parameters. Correlation coefficient can be used to determine if a climatic element is positively or adversely associated with the increase of dengue fever cases (35). Various correlation strategies have been utilized in various research, and they have proven to help remove the climatic factors that were not very significant. The Pearson's correlation coefficient determines the statistical relationship between two variables. It is based on the covariance method. It is widely acknowledged as the most effective approach for determining the connection between two variables of interest. It offers information on the amount and direction of the link, or correlation, between the two variables. A non-parametric test called Spearman rank correlation predicts the degree of relationship among two variables. Pearson Correlation Coefficient (36) and Spearman's Rank Correlation Test (37) have been widely used. Sourabh Bal et al. carried out their study of dengue occurrence based on climatic factors for the region Kolkata, India, using auto-correlation coefficient and partial auto-correlation coefficient values. An auto-correlation analysis is used to examine dengue cases affected by prior instances. In addition, the Pearson correlation was used to assess for collinearity between the various climate variables (38). In the study carried out by Daniel Sánchez-Hernández et al. carried out correlation analysis to inspect they built the relationships of proximity, environmental, social factors and location about the occurrence of dengue, a multivariable logistic regression model. A logistic curve may be produced by graphing the connection between the explanatory and predicted variables, which will help us evaluate the correlation strength (39). Various correlation techniques have been according to the review Pearson, and spearman's correlation test were widely used and effective in most of the studies.

3.5. Dengue Prediction Modeling Techniques

Various models can predict dengue incidence based on climatic variables (40). Several studies carried out have used different models are described along with the results. **Figure 8** depicts the steps and techniques that will required and are essential to develop an early monitoring dengue surveillance system.

3.5.1. Support Vector Regression

SVMs and Support Vector Regression are both based on similar concepts. SVR's main goal is to find the best-fitting line. The hyperplane with the greatest number of points is the best fit line in SVR. In SVR, unlike other regression models, the goal is to fit the best line inside a threshold value. The complexity of SVR's reasonable time grows more than quadratically with the number of samples, making it difficult to scale to datasets with more than a few tens of thousands of samples. Linear SVR is faster than SVR; however, it solely considers the linear kernel. As the cost function ignores samples whose prediction is close to their goal. Based on the evaluation metrics like RMSE and MAE, SVR with linear kernel proved to be quite good at forecasting the number of dengue incidents in Jakarta. Their experiments with various penalty parameters for SVR with linear kernel yielded quite accurate outcomes. When cross-correlation between variables,

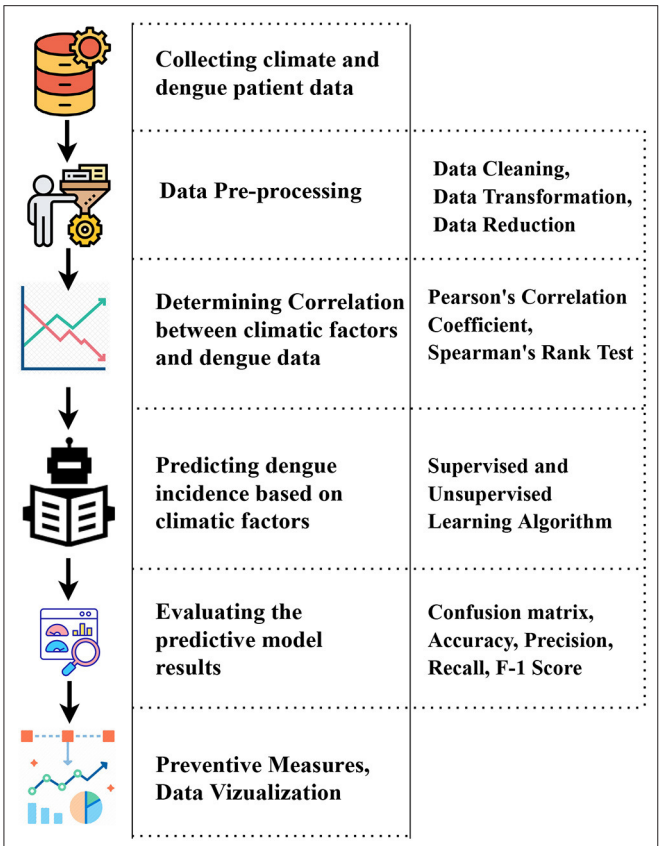


FIGURE 8 | Process of developing a dengue surveillance system.

the linear kernel has a higher prediction accuracy than the radial kernel (41).

3.5.2. Random Forest

Random Forest (42) is an ensemble technique that builds many separate bootstrapped trees from random small subsets of the data using bootstrap aggregation (bagging) (43). As it can handle large numbers of target variables even in complicated interactions, RF is utilized to tackle issues in classification and regression (44). Every tree in the random forest produces a class prediction, and the class with the most votes becomes the prediction of our model. Before training, three hyperparameters of random forest algorithms have to be established. The factors to be considered are the size of nodes, the number of trees and the number of characteristics sampled. This technique is used by Micanaldo et al. (45) in predicting dengue based on various explanatory variables.

3.5.3. Model-Based Recursive Partitioning

The combined effects of the chosen explanatory variables on OI and Dengue fever incidence, Micanaldo et al. (45) used a Model-Based (MOB) recursive partitioning. MOB is similar to CART (classification and regression tree) methods, which recursively partition datasets into subsets depending on independent variables at each step (46). The following steps are repeated iteratively by the MOB algorithm: (1) fit the data to a user-defined linear regression equation; (2) determine if other variables have an impact on model parameters.; (3) if so, using a threshold that results in the most significant changes in the linear model parameters based on the M-fluctuation test, split the model and data into two groups concerning the covariate.; and (4) In each of the subsamples, repeat steps 1–3. Until a specific stop condition is satisfied, the steps are repeated.

3.5.4. Distributed Lag Non-linear Models

Distributed Lag Non-Linear Model is modeling framework for representing links in time series data with non-linear and delayed effects flexibly. The establishment of cross basis, bi-dimensional functional space created by the combination of 2 sets of basis functions that characterize the relationships in the predictor and lag dimensions, respectively, is the cornerstone of this technique.

3.5.5. Generalized Linear Model

Generalized Linear Model is a sophisticated statistical modeling technique that enables us to construct a linear relationship between the answer and the predictors, even though their underlying relationship is not linear. This is possible because of the employment of a link function, which connects the response variable to a linear model. Here, the error distribution of the response variable does not have to be regular.

3.5.6. Generalized Additive Model

Generalized Additive Models are a sort of statistical model. Many nonlinear smooth functions replace the typical linear relationship between the response and targets to represent and capture data non-linearities. These flexible and soft approaches allow us to fit linear models that are either linearly or non-linearly dependent on many predictors to capture nonlinear correlations between response and predictors.

3.5.7. Multilayer Perceptron

A feed-forward ANN called a Multilayer Perceptron is a feed-forward artificial neural network. A hidden layer, an input layer, and an output layer are at least three levels of nodes. For a long time, ANN has been a reliable perceptive classifier for various applications, including medical diagnosis and disease early detection. MLP adapts the classic linear perceptron and employs a supervised learning technique to propagate the network. As a result, it can distinguish facts that cannot be separated. A perceptron forms a linear combination using input weights to produce a single output based on numerous real-valued inputs. They train on a collection of input-output pairs to learn how to express the correlation between inputs and outputs. The model's parameters, or weights and biases, are adjusted throughout training to minimize inaccuracy. Back-propagation is used to change the consequences and preferences about the error, which can be quantified in various ways (47).

3.5.8. K Nearest Neighbor Regression

Scavuzzo et al. (48) used the K-Neighbors Regressor module. Based on k-nearest neighbors, this technique infers a regression. Local interpolation of the training set's targets

in the neighborhood is used to forecast the target. Only the first five principal components are used to deconstruct the original data. Four neighbors, Chebyshev metric, brute force and uniform weight, were the tuning settings available. Out of all the tested models, this model gave beast results for modeling the dengue vector population.

3.5.9. Auto-Regressive Integrated Moving Average Model

Xavier et al. (15) chose the ARIMA model, a family of autoregressive moving averages, to describe the relationship among the count of positive dengue cases (dependent variable) and meteorological parameters (explanatory variables). The ARIMA model's primary goal is to directly simulate the autocorrelation in a time series to capture it. Solid underlying mathematics and statistical theory are commonly used to predict time series data, making it easier to construct expected ranges. The ARIMA model is very adaptable, capturing a wide range of patterns. Choosing order and differencing are two fundamental principles in the ARIMA model.

3.5.10. Time Series Poisson Regression Model

Sang et al. (56) to determine the relationship between meteorological conditions and local dengue count, researchers developed this model called Time Series Poisson Regression. According to this model, dengue count was positively related to dengue count in the preceding month, imported cases in the previous month, the minimum temperature during the last month, and accumulative precipitation with 3-month lags.

Table 2 illustrates the detailed representation of latest research papers which has undertaken the study to understand the relation between dengue and meteorological patterns and the need to develop a dengue surveillance system.

4. DISCUSSION

According to the number of people infected, dengue fever is the most common arboviral disease in the world. Understanding the precise relationship between meteorology and dengue transmission is not a simple process because dengue transmission involves dengue viruses, vectors, and susceptible people. Furthermore, forecasting the future of dengue under various climate change scenarios involves a complete understanding of the association between climate and dengue and future climate and other variables. Despite this, a lot of progress has been achieved in this area. Despite the advances achieved in forecasting dengue's future, numerous uncertainties remain (59). To begin with, socio-demographic factors play a significant influence in dengue transmission, and including socio-demographic components into future dengue, projections remains a challenge. Second, in previous studies estimating the future of dengue, the increasing temperature has been commonly employed as a climate change indicator, with rainfall and humidity being under-researched. According to Hales et al., the climate indicator that most correctly predicts the occurrence of dengue fever is vapor pressure, which is a temperature and humidity measurements. However, the relationships between numerous environmental conditions and dengue transmission are complex and sometimes non-linear. Dengue modeling is an essential technique for early detection of dengue outbreaks, assessing risk factors for Severe Dengue, and possibly controlling the disease's vectors. Although much research has been done on these topics, it is critical to understand what areas of dengue modeling still have to be explored to build future research that will significantly reduce disease morbidity rates. The primary goal of this study was to provide an overview of dengue modeling and identify critical issues for future research. This section focuses on the limitations of the studies that have been reviewed. Some research problems or opportunities are described based on those restrictions.

4.1. Limitations, Research Challenges and Future Work Directions

4.1.1. Inclusion of Micro Climatic Factors

Researchers have widely used temperature, humidity, and rainfall are macro climatic factors. But to bring in more explanatory variables that can explain the rise in dengue incidence rise can aid in building a better dengue surveillance system (32, 55, 60, 61).

SOLUTION: There are many climatic variables like wind speed, air pressure, vegetation index, and we can add many other microclimatic factors to the model for better dengue incidence.

4.1.2. Exploring Different Geographic Locations

Many types of research stated the limitation that how dengue outbreak prediction model developed for one region cannot be used accurately for different areas as dengue incidence may or may not be related to the same climatic factors of one place and due to weather fluctuations and pattern of coastal regions might not give excellent predictive results from the model that was developed for non-coastal regions (55, 62–64).

SOLUTION: Exploring different regions other than that have already been investigated or studying the weather patterns of the country where more minor number studies have been carried out in predicting dengue incidence based on the weather forecast can help in understanding the relationship between different patterns of weather for various regions and dengue rise and eventually would help build a dengue surveillance system.

4.1.3. Data Limitations

Obtaining a dengue dataset that is completely fit for the model and has all characteristics is difficult to find. There are many issues faced by the researchers in this area like i) data being incomplete, ii) data being not available in required time lag like weekly or monthly and for a specific time-period and iii) data that is available only on a macro level like country-level or state-level and not on micro-level like city-level (15, 28, 45, 54, 63, 65).

SOLUTION: Data cleaning can be handled by using various data pre-processing techniques based on the data type. When data is not completed available, taking the average or mean of the available information can be helpful. To obtain data at the micro-level can be a tedious task, but with all the proper procedures, the patient's data can be obtained from the Health Department of the respective region.

TABLE 2 | Detailed representations of different models and feature engineering techniques used in different studies.

References	Year	Region	Techniques	Contributions
Leandro et al. (15)	2021	Brazil	Feature Engineering: ARMAX model	ARMAX model best fits the data used in this study and produced the dengue incidence count with good precision for future.
Joseph et al. (27)	2021	Philippines	Feature Engineering: Pearson's Coefficient Predictive Model: Linear Regression, Exponential Regression	Regression modeling estimated the annual dengue force of infection across urban centers from the age of those with infections and the transmission intensity showed significant spatiotemporal variation.
Daniel et al. (39)	2019	Mexico	Feature Engineering: Multivariate Analysis Predictive Model: Multivariable Logistic Regression Model-MLRM	The connection between dengue and explanatory factors was evaluated via MLRM. A high spatial resolution map was created to highlight the most likely patterns of dengue risk.
Sourabh et al. (38)	2020	India	Feature Engineering: AutoCorrelation Coefficient, Partial Correlation Coefficient and Cross-correlation coefficient Predictive Model: Poisson Distribution Regression Model and Zero Inflated Poisson Regression Model-ZIP	Based on the numerous explanatory factors, the ZIP Model was used to predict the severity of dengue fever in Kolkata.
Oladimeji et al. (34)	2021	Brazil	Feature Engineering: Statistical Analysis Predictive Model: LSTM, RNN	This study used RNN to forecast dengue count and it also uses clustering before modeling which aggregates dengue count with similar temporal patterns.
Micanaldo et al. (45)	2021	Phillipines	Feature Engineering: Cross-correlation Analysis, Variable selection using Random Forest algorithm Predictive Model: Model-based recursive partitioning	MOB recursive partitioning displayed high correlation between dengue transmission and climatic factors and gave accurate predictions.
Sandali et al. (57)	2021	India	Feature Engineering: Data Imputation and Normalization, filling missing values with mean Predictive Model: Artificial Neural Network	A new ANN based multimodal outbreak prediction algorithm is used to predict dengue with the accuracy of 86 percent.
Ivan et al. (41)	2020	Indonesia	Feature Engineering: Cross-correlation, Augmented Dickey Fuller Test Predictive Model: Support Vector Regression	SVR with a linear kernel was applied to climate and dengue incidence data for predicting dengue count and this study also provides a comparative analysis of linear and radial kernel.
Mohd et al. (58)	2019	Malaysia	Feature Engineering: Average Nearest Neighbor (ANN) Predictive Modeling: Spatial Clustering	The hotspot locations were detected using ANN and Kernel density estimation. Based on past data, this study found that it is feasible to estimate dengue risk.
Sabrina et al. (28)	2021	Bangladesh	Feature Engineering: Statistical Analysis Predictive Model: Generalized Additive Model, Generalized Linear Model	The GLM and GAM models were used to show the link between dengue and environmental variables.
Teerawad et al. (32)	2021	Thailand	Feature Engineering: Spearman's Rank Correlation Test Predictive Model: Poisson Regression Model and ARIMA model	Spatial and Temporal modeling of dengue fever transmission was presented and Poisson regression model was used for prediction of dengue based on various climatic factors
Gayan et al. (54)	2018	Sri Lanka	Feature Engineering: Pearson Correlation Coefficient, AutoCorrelation Coefficient, Partial Correlation Coefficient Predictive Model: Time Series Regression Model	The suggested weather-based forecasting algorithm provides high-precision warnings of oncoming dengue outbreaks and epidemics up to one month ahead of time.
Felestin et al. (55)	2021	Malaysia	Feature Engineering: Pearson Correlation Coefficient Predictive Model: Bayes network (BN) models, support vector machine (SVM), RBF tree, decision table and Naive Bayes	The TempeRain factor (TRF), a novel risk factor, was discovered and employed as an input parameter for a dengue epidemic prediction model. The Bayes network produced reliable findings.
Rachel et al. (33)	2021	Brazil	Feature Engineering: Pearson Correlation Test Predictive Model: Spatiotemporal Bayesian Hierarchical Model, Distributed Lag Non-Liner Model	Space-varying, non-linear, and delayed connections between hydrometeorological parameters and dengue incidence are described using coupled spatiotemporal Bayesian hierarchical models with distributed lag non-linear models.
Wu et al. (53)	2021	China	Feature Engineering: Spearman's Correlation Coefficient Predictive Model: Ecological Niche Models(ENM)	ENMs determined the non-random association between dengue count and meteorological factors. Maxnet model was used to predict dengue incidence.

4.1.4. Considering Vector Density and Mosquito Larvae Data

The main issue highlighted in most of the studies was the correlation between climatic factors and dengue incidence; whereas those are not the only factors that contribute to increased disease transmission of dengue, the main factor is the vector that carries the transmission (32, 33, 49).

SOLUTION: Considering the vector density, mosquito larvae data or land use would aid in gaining a better understanding of the main factor behind the rise in dengue cases. Considering

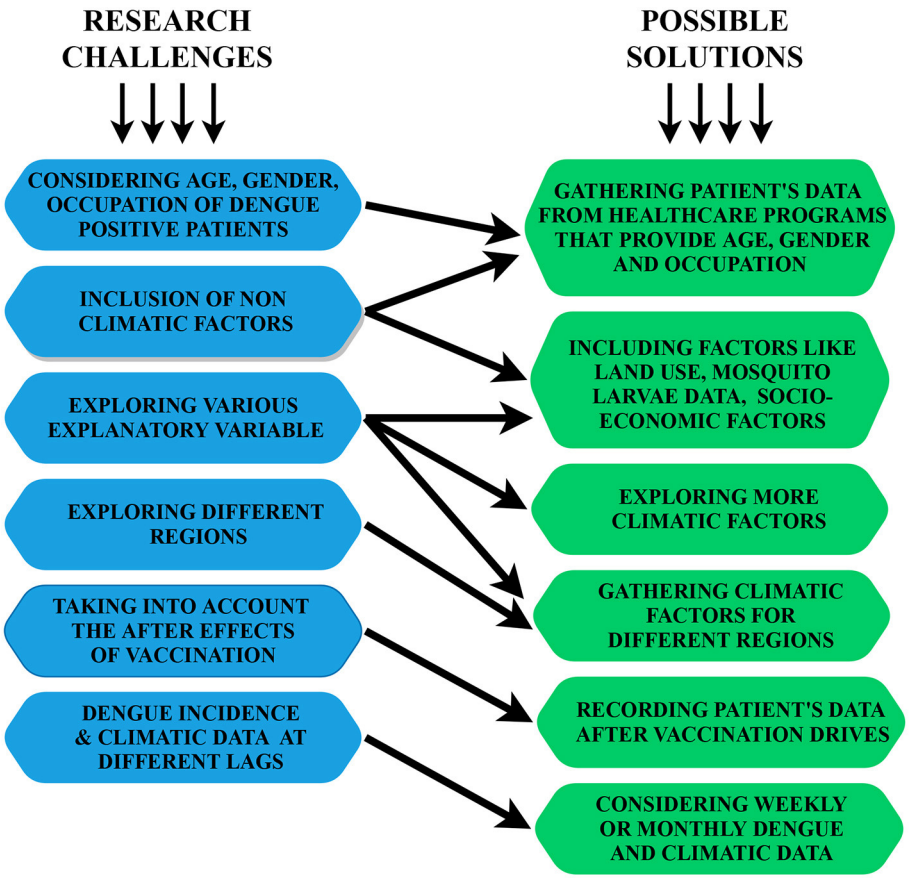


FIGURE 9 | A logical mapping of research challenges and possible solutions.

TABLE 3 | Research challenges.

Reference	Year	Research challenge	Discussion
Nuraini et al. (64)	2021	Stratifying dengue incidences based on age, gender and occupation	Classifying the population based on different vulnerable groups can help understand which crowd has been affected the most and make future predictions accordingly
Nuraini et al. (64)	2021	Different Regions	Climatic and non-climatic factors vary in different regions. Hence, it is important to understand the geographic and weather patterns to predict dengue accurately for a particular region.
Sriklin et al. (32)	2021	Stratifying dengue incidences based on age, gender and occupation	Non atmospheric data can help understand the factors contributing to dengue and help prevent dengue at an early stage.
Nuraini et al. (64)	2021	Climatic Factors	Rise in dengue incidences became correlated to the temperature and rainfall which would help make dengue outbreak predictions based on weather forecast.
Sriklin et al. (32)	2021	Dataset	Available datasets of dengue incidence were mostly yearly or monthly, hence weekly data could help make accurate predictions. And also the dengue incidences dataset were not accurate and had some missing data regarding patients.

these parameters would help predict the outbreak better and take preventive measures well in advance.

4.1.5. Taking Into Account the Effect of Vaccination After the Outbreak

The studies carried until now focus on determining the correlation between the climatic factor and dengue incidence and then using weather forecast of coming time to predict the dengue incidence and taking preventive measures. Still, the shortcoming that few researchers found is that there might be a fluctuation in predicting dengue incidence when every infected person might be vaccinated (66–68).

SOLUTION: As and when the vaccination of patients is done, the data of the patient's immune can be determined and kept a

record of. Further, that data can be used whenever the prediction for other outbreaks is being carried out. **Figure 9** represents the logical mapping of limitations and their possible solutions.

4.2. Contributions

Dengue Surveillance system can aid the health sector in taking preventive measures for dengue outbreaks in advance. Also, it can help to be prepared for the worst of situations and reduce the vulnerability of epidemics. Several research studies have predicted the epidemic using machine learning approaches, including climatic factors. The present review discusses the most common data sources, data pre-processing stages, most common correlation techniques for dengue incidence and climatic factors, machine learning techniques, limitations, and future research directions. **Table 3** represents the research challenges of the current study. The contribution of the presented student are:

- The present study gives a detailed insight into the amount of work done in predicting dengue incidence using different machine learning techniques based on various explanatory variables.
- To assist readers, we have discussed several data sources for dengue incidence count and weather variables.
- different approaches for determining the relation between dengue incidence and explanatory variables and various explanatory variables (climatic factors, demographic variables, vulnerable groups).
- Furthermore, we have discussed various dengue modeling techniques for predicting outbreaks.
- This work's main contribution is to outline open research concerns and limitations of various studies.

5. CONCLUSION AND FUTURE ENHANCEMENTS

A systematic literature review was conducted on predicting dengue based on climatic change using machine learning techniques. The main objective was to understand the research and studies that have been carried in building a dengue outbreak prediction model. Forty articles were chosen and analyzed to determine the state-of-the-art from many scientific libraries. The results of the review represent that dengue modeling is continuously growing. Logistic Regression, LASSO Regression, and Support Vector Machine approaches were the most commonly used diagnostic models. Because of their ease of implementation and interpretation, they are the most widely used modeling techniques. Although alternative strategies, such as decision trees, are simple to understand, they have a large number of nodes and take a significant amount of mental work to comprehend a specific forecast. These models, on the other hand, are just a set of coefficients, which makes it appealing to learn about the impact of attributes on the predictor variable. Furthermore, continuous independent variables do not have a normal distribution, and continuous and discrete predictors can be used in the regression. In terms of feature types, climate data was the most commonly used in these models. Dengue incidence data of patients is not readily available through APIs, which has been a restriction in many studies and should be considered for future work by forecasting dengue incidence for sensitive groups such as the age and gender of patients. The examined articles had several flaws, including the lack of documentation of the pre-processing procedures employed, the unavailability of patient data, and incomplete data. Following the review of the papers' strengths and shortcomings, future research projects were identified: i) using microclimatic parameters such as wind speed, air pressure, or vegetation index, ii) using demographic and socioeconomic aspects, iii) exploring climatic conditions in different places, iv) using vector density data, and v) taking vaccination drives into account when building a prediction model. Forecasting the future of dengue fever in the context of climate change can assist governments and public health professionals in implementing timely and preventative steps to protect people from dengue in the future. To sum up, climate change is creating an alarming situation affecting many sectors. Dengue has been a reason of concern for a long time. Hence, monitoring the fluctuations of weather patterns can aid in creating a dengue surveillance system that would be of great help to the health sector in taking preventive measures well in advance.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SPat, SPan, and DB: conceptualization. SB, QI, and SI: funding acquisition. DB and SPat: investigation and methodology, writing of the original draft, and data curation. SB: project administration and visualization. QI and SI: resources and supervision. SPan and SB: writing of the review and editing. SPan: validation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project no. GRANT 268).

REFERENCES

1. Parashar G, Chaudhary A, Rana A. Systematic mapping study of AI/machine learning in healthcare and future directions. *SN Comput Sci*. (2021) 2:1–8. doi: 10.1007/s42979-021-00848-6
2. Raisch S, Krakowski S. Artificial intelligence and management: the automation-augmentation paradox. *Acad Manag Rev*. (2021) 46:192–210. doi: 10.5465/amr.2018.0072
3. Galaz V, Centeno MA, Callahan PW, Causevic A, Patterson T, Brass I, et al. Artificial intelligence, systemic risks, and sustainability. *Technol Soc*. (2021) 67:101741. doi: 10.1016/j.techsoc.2021.101741

4. Dixit P, Payal M, Dutt V, Tuteja A. A review of machine learning approaches in clinical healthcare. *Intell Healthcare*. (2021) 113:243–58. doi: 10.1007/978-3-030-67051-1_15
5. Sunarti S, Rahman FF, Naufal M, Risky M, Febriyanto K, Masnina R. Artificial intelligence in healthcare: opportunities and risk for future. *Gaceta Sanitaria*. (2021) 35:S67–70. doi: 10.1016/j.gaceta.2020.12.019
6. Kiener M. Artificial intelligence in medicine and the disclosure of risks. *Ai & Society*. (2021) 36:705–13. doi: 10.1007/s00146-020-01085-w
7. Hannah L. *Climate Change Biology*. Academic Press (2021).
8. Iwendi C, Huescas C, Chakraborty C, Mohan S. COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients. *J Exp Theor Artif Intell*. (2022) 1–21. doi: 10.1080/0952813X.2022.2058097
9. Nations U. *What Is Climate Change*? United Nations. Available online at: https://www.un.org/en/climatechange/what-is-climate-change
10. CDC. *Climate Effects on Health*. CDC (2021). Available online at: https://www.cdc.gov/climateandhealth/effects/default.htm.
11. Ebi KL, Nealon J. Dengue in a changing climate. *Environ Res*. (2016) 151:115–23. doi: 10.1016/j.envres.2016.07.026
12. Christophrs SS. Aëdes aegyptì (L.) the yellow fever mosquito; its life history, bionomics and structure. Aëdes aegyptì (L) the yellow fever mosquito; its life history, bionomics and structure (1960).
13. Wilke ABB, Vasquez C, Carvajal A, Medina J, Chase C, Cardenas G, et al. Proliferation of Aedes aegypti in urban environments mediated by the availability of key aquatic habitats. *Sci Rep*. (2020) 10:1–10. doi: 10.1038/s41598-020-69759-5
14. WHO. *Dengue and Severe Dengue*. Available online at: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.
15. Xavier LL, Honório NA, Pessanha JFM, Peiter PC. Analysis of climate factors and dengue incidence in the metropolitan region of Rio de Janeiro, Brazil. *PLoS ONE*. (2021) 16:e0251403. doi: 10.1371/journal.pone.0251403
16. Duarte JL, Diaz-Quijano FA, Batista AC, Giatti LL. Climatic variables associated with dengue incidence in a city of the Western Brazilian Amazon region. *Rev Soc Bras Med Trop*. (2019) 52:e20180429. doi: 10.1590/0037-8682-0429-2018
17. Kakarla SG, Caminade C, Mutheneni SR, Morse AP, Upadhyayula SM, Kadiri MR, et al. Lag effect of climatic variables on dengue burden in India. *Epidemiol Infect*. (2019) 147:e170. doi: 10.1017/S0950268819000608
18. Arcari P, Tapper N, Pfueller S. Regional variability in relationships between climate and dengue/DHF in Indonesia. *Singap J Trop Geogr*. (2007) 28:251–72. doi: 10.1111/j.1467-9493.2007.00300.x
19. Ong YY, Tan GE. *Climate Variability and Dengue in Singapore, Fiji, and Hong Kong: Small Bite, Big Threat*. London: SAGE Publications Sage UK.
20. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Climatic factors affecting dengue Haemorrhagic fever incidence in southern Thailand (2005).
21. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. (2013) 496:504–7. doi: 10.1038/nature12060
22. Climate change and health Available online at: https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health.
23. Li C, Wu X, Sheridan S, Lee J, Wang X, Yin J, et al. Interaction of climate and socio-ecological environment drives the dengue outbreak in epidemic region of China. *PLoS Negl Trop Dis*. (2021) 15:e0009761. doi: 10.1371/journal.pntd.0009761
24. Wu PC, Guo HR, Lung SC, Lin CY, Su HJ. Weather as an effective predictor for occurrence of dengue fever in Taiwan. *Acta Trop*. (2007) 103:50–57. doi: 10.1016/j.actatropica.2007.05.014
25. Wong PF, Wong LP, AbuBakar S. Diagnosis of severe dengue: challenges, needs and opportunities. *J Infect Public Health*. (2020) 13:193–8. doi: 10.1016/j.jiph.2019.07.012
26. Laboratory Diagnosis and Diagnostic Tests - Dengue - NCBI Bookshelf. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK143156/
27. Biggs JR, Sy AK, Sherratt K, Brady OJ, Kucharski AJ, Funk S, et al. Estimating the annual dengue force of infection from the age of reporting primary infections across urban centres in endemic countries. *BMC Med*. (2021) 19:217. doi: 10.1186/s12916-021-02101-6
28. Islam S, Haque CE, Hossain S, Hanesiak J. Climate variability, dengue vector abundance and dengue fever cases in dhaka, bangladesh: a time-series study. *Atmosphere*. (2021) 12:905. doi: 10.3390/atmos12070905
29. Jannat-Khah DP, Unterbrink M, McNairy M, Pierre S, Fitzgerald DW, Pape J, et al. Treating loss-to-follow-up as a missing data problem: a case study using a longitudinal cohort of HIV-infected patients in Haiti. *BMC Public Health*. (2018) 18:1–11. doi: 10.1186/s12889-018-6115-0
30. Mello-Román JD, Mello-Román JC, Gómez-Guerrero S, García-Torres M. Predictive models for the medical diagnosis of dengue: a case study in paraguay. *Comput Math Methods Med*. (2019) 2019:7307803. doi: 10.1155/2019/7307803
31. Alkhaldy I. Modelling the association of dengue fever cases with temperature and relative humidity in Jeddah, Saudi Arabi'a generalised linear model with break-point analysis. *Acta Trop*. (2017) 168:9–15. doi: 10.1016/j.actatropica.2016.12.034
32. Sriklin T, Kajornkasirat S, Puttinaovarat S. Dengue transmission mapping with weather-based predictive model in three southernmost provinces of thailand. *Sustainability*. (2021) 13:6754. doi: 10.3390/su13126754
33. Lowe R, Lee SA, O'Reilly KM, Brady OJ, Bastos L, Carrasco-Escobar G, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary health*. (2021) 5:e209–19. doi: 10.1016/S2542-5196(20)30292-8
34. Mudele O, Frery AC, Zanandrez LFR, Eiras AE, Gamba P. Dengue vector population forecasting using multisource earth observation products and recurrent neural networks. *IEEE J Select Top Appl Earth Observat Remote Sens*. (2021) 14:4390–404. doi: 10.1109/JSTARS.2021.3073351
35. Islam MZ, Rutherford S, Dung Phung M, Uzzaman N, Baum S, Huda MM, et al. Correlates of climate variability and dengue fever in two metropolitan cities in bangladesh. *Cureus*. (2018) 10:e3398. doi: 10.7759/cureus.3398
36. Pearson Correlation Coefficient.
37. Zar JH. Significance testing of the spearman rank correlation coefficient. *J Am Stat Assoc*. (1972) 67:578–80. doi: 10.1080/01621459.1972.10481251
38. Bal S, Sodoudi S. Modeling and prediction of dengue occurrences in Kolkata, India, based on climate factors. *Int J Biometeorol*. (2020) 64:1379–91. doi: 10.1007/s00484-020-01918-9
39. Sánchez-Hernández D, Aguirre-Salado CA, Sánchez-Díaz G, Aguirre-Salado AI, Soubervielle-Montalvo C, Reyes-Cárdenas O, et al. Modeling spatial pattern of dengue in North Central Mexico using survey data and logistic regression. *Int J Environ Health Res*. (2019) 31:872–88. doi: 10.1080/09603123.2019.1700938
40. Pley C, Evans M, Lowe R, Montgomery H, Yacoub S. Digital and technological innovation in vector-borne disease surveillance to predict, detect, and control climate-driven outbreaks. *Lancet Planetary Health*. (2021) 5:e739–45. doi: 10.1016/S2542-5196(21)00141-8
41. Tanawi IN, Vito V, Sarwinda D, Tasman H, Hertono GF. support vector regression for predicting the number of dengue incidents in DKI Jakarta. *Procedia Comput Sci*. (2021) 179:747–53. doi: 10.1016/j.procs.2021.01.063
42. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
43. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
44. Garge NR, Bobashev G, Eggleston B. Random forest methodology for model-based recursive partitioning: the mobForest package for R. *BMC Bioinformatics*. (2013) 14:1–8. doi: 10.1186/1471-2105-14-125
45. Micanaldo FE, Carvajal TM, Ryo M, Nukazawa K, Amalin DM, Watanabe K. Dengue disease dynamics are modulated by the combined influences of precipitation and landscape: a machine learning approach. *Sci Total Environ*. (2021) 792:148406. doi: 10.1016/j.scitotenv.2021.148406
46. Pirkle CM, Wu YY, Zunzunegui MV, Gómez JF. Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components: findings from the International Mobility in Aging Study. *BMJ Open*. (2018) 8:e018680. doi: 10.1136/bmjopen-2017-018680
47. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst*. (2021) 1–15. doi: 10.1007/s00530-021-00774-w

48. Scavuzzo JM, Trucco F, Espinosa M, Tauro CB, Abril M, Scavuzzo CM, et al. Modeling Dengue vector population using remotely sensed data and machine learning. *Acta Trop*. (2018) 185:167–75. doi: 10.1016/j.actatropica.2018.05.003
49. Cheng J, Bambrick H, Frentiu FD, Devine G, Yakob L, Xu Z, et al. Extreme weather events and dengue outbreaks in Guangzhou, China: a time-series quasi-binomial distributed lag non-linear model. *Int J Biometeorol*. (2021) 65:1033–42. doi: 10.1007/s00484-021-02085-1
50. Meng H, Xiao J, Liu T, Zhu Z, Gong D, Kang M, et al. The impacts of precipitation patterns on dengue epidemics in Guangzhou city. *Int J Biometeorol*. (2021) 65:1929–37. doi: 10.1007/s00484-021-02149-2
51. Shashvat K, Basu R, Bhondekar AP. Application of time series methods for dengue cases in North India (Chandigarh). *J Public Health*. (2019) 29:433–41. doi: 10.1007/s10389-019-01136-7
52. Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: a case study in China. *PLoS Negl Trop Dis*. (2017) 11:e0005973. doi: 10.1371/journal.pntd.0005973
53. Wu W, Ren H, Lu L. Increasingly expanded future risk of dengue fever in the Pearl River Delta, China. *PLoS Negl Trop Dis*. (2021) 15:e0009745. doi: 10.1371/journal.pntd.0009745
54. Withanage GP, Viswakula SD, Nilmini Silva Gunawardena YI, Hapugoda MD. A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasites Vectors*. (2018) 11:1–10. doi: 10.1186/s13071-018-2828-2
55. Yavari Nejad F, Varathan KD. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. *BMC Med Inform Decis Mak*. (2021) 21:141. doi: 10.1186/s12911-021-01493-y
56. Sang S, Gu S, Bi P, Yang W, Yang Z, Xu L, et al. Predicting unprecedented dengue outbreak using imported cases and climatic factors in Guangzhou, 2014. *PLoS Negl Trop Dis*. (2015) 9:e0003808. doi: 10.1371/journal.pntd.0003808
57. Raizada S, Mala S, Shankar A. *Vector-Borne Disease Outbreak Prediction Using Machine Learning Techniques*. In: Prakash KB, Kannan R, Alexander SA, Kanagachidambaresan GR, editors. Cham: Springer International Publishing (2021). p. 227–41.
58. Yajid MZM, Dom NC, Camalxaman SN, Nasir RA. Spatial-temporal analysis for identification of dengue risk area in Melaka Tengah district. *Geocarto Int*. (2019) 35:1570–9. doi: 10.1080/10106049.2019.1581265
59. Shabbir A, Shabbir M, Javed AR, Rizwan M, Iwendi C, Chakraborty C. Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals. *J Exp Theor Artif Intell*. (2022) p. 1-28. doi: 10.1080/0952813X.2021.1960634
60. Lee HS, Nguyen-Viet H, Nam VS, Lee M, Won S, Duc PP, et al. Seasonal patterns of dengue fever and associated climate factors in 4 provinces in Vietnam from 1994 to (2013). *BMC Infect Dis*. (2017) 17:1–9. doi: 10.1186/s12879-017-2326-8
61. Erraguntla M, Dave D, Zapletal J, Myles K, Adelman ZN, Pohlenz TD, et al. Predictive model for microclimatic temperature and its use in mosquito population modeling. *Sci Rep*. (2021) 11:1–11. doi: 10.1038/s41598-021-98316-x
62. Wu CH, Kao SC. Knowledge discovery in open data for epidemic disease prediction. *Health Policy Technol*. (2021) 10:126–34. doi: 10.1016/j.hlpt.2021.01.001
63. Espinoza-Gomez F, Newton-Sanchez OA, Nava-Zavala AH, Zavala-Cerna MG, Rojas-Larios F, Delgado-Enciso I, et al. Demographic and climatic factors associated with dengue prevalence in a hyperendemic zone in Mexico: an empirical approach. *Trans R Soc Trop Med Hyg*. (2021) 115:63–73. doi: 10.1093/trstmh/traa083
64. Nuraini N, Fauzi IS, Fakhruddin M, Sopaheluwakan A, Soewono E. Climate-based dengue model in Semarang, Indonesia: Predictions and descriptive analysis. *Infect Dis Model*. (2021) 6:598–611. doi: 10.1016/j.idm.2021.03.005
65. Gao Q, Liu Z, Xiang J, Tong M, Zhang Y, Wang S, et al. Forecast and early warning of hand, foot, and mouth disease based on meteorological factors: Evidence from a multicity study of 11 meteorological geographical divisions in mainland China. *Environ Res*. (2021) 192:110301. doi: 10.1016/j.envres.2020.110301
66. Chen Y, Ong JHY, Rajarethinam J, Yap G, Ng LC, Cook AR. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Med*. (2018) 16:129. doi: 10.1186/s12916-018-1108-5
67. Colón-González FJ, Fezzi C, Lake IR, Hunter PR. The effects of weather and climate change on dengue. *PLoS Negl Trop Dis*. (2013) 7:e2503. doi: 10.1371/journal.pntd.0002503
68. World Health Organization. Comprehensive guideline for prevention and control of dengue and dengue haemorrhagic fever (2011).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bhatia, Bansal, Patil, Pandya, Ilyas and Imran. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fpubh.2022.898355

Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired

Surbhi Bhatia[1]*, *Ajantha Devi*[2]*, *Razan Ibrahim Alsuwailem*[1] and *Arwa Mashat*[3]

[1] Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Al Hasa, Saudi Arabia, [2] Research Head, AP3 Solutions, Chennai, India, [3] Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, Saudi Arabia

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Pradeep Singh,
KIET Group of Institutions, India
Zaki Khan,
University of Worcester,
United Kingdom
Praise Young,
Nnamdi Azikiwe University, Nigeria
Ayodeji Olateju,
University of Bolton, United Kingdom

***Correspondence:**
Surbhi Bhatia
sbhatia@kfu.edu.sa
Ajantha Devi
research_head@ap3-solutions.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 17 March 2022
Accepted: 25 April 2022
Published: 27 May 2022

Citation:
Bhatia S, Devi A, Alsuwailem RI and Mashat A (2022) Convolutional Neural Network Based Real Time Arabic Speech Recognition to Arabic Braille for Hearing and Visually Impaired. Front. Public Health 10:898355. doi: 10.3389/fpubh.2022.898355

Natural Language Processing (NLP) is a group of theoretically inspired computer structures for analyzing and modeling clearly going on texts at one or extra degrees of linguistic evaluation to acquire human-like language processing for quite a few activities and applications. Hearing and visually impaired people are unable to see entirely or have very low vision, as well as being unable to hear completely or having a hard time hearing. It is difficult to get information since both hearing and vision, which are crucial organs for receiving information, are harmed. Hearing and visually impaired people are considered to have a substantial information deficit, as opposed to people who just have one handicap, such as blindness or deafness. Visually and hearing-impaired people who are unable to communicate with the outside world may experience emotional loneliness, which can lead to stress and, in extreme cases, serious mental illness. As a result, overcoming information handicap is a critical issue for visually and hearing-impaired people who want to live active, independent lives in society. The major objective of this study is to recognize Arabic speech in real time and convert it to Arabic text using Convolutional Neural Network-based algorithms before saving it to an SD card. The Arabic text is then translated into Arabic Braille characters, which are then used to control the Braille pattern via a Braille display with a solenoid drive. The Braille lettering triggered on the finger was deciphered by visually and hearing challenged participants who were proficient in Braille reading. The CNN, in combination with the ReLU model learning parameters, is fine-tuned for optimization, resulting in a model training accuracy of 90%. The tuned parameters model's testing results show that adding the ReLU activation function to the CNN model improves recognition accuracy by 84 % when speaking Arabic digits.

Keywords: braille, Speech to Text, speech to Braille, speech recognition, text to braille, Arabic digit

INTRODUCTION

Louis Braille invented Braille, a worldwide method of reading and writing for visually impaired people, in 1821. The Braille cell is the system's smallest unit. Because of the prosody (intonation) of the system, visually and hearing-impaired individuals who are skilled in Finger Braille can communicate words and express a variety of emotions (1). Because there are so few people who

can read Finger Braille, visually and hearing-impaired people can only communicate through an interpreter. As a result, those who are visually and hearing impaired are severely limited in their participation. Finger Braille input techniques, including a wearable device, have been developed (2). Created a wearable input device employing accelerometers affixed on the tops of rings and invented finger Braille gloves with accelerometers integrated in the fingers (3). In addition (4), created a wristband with an eighteen-accelerometer Finger Braille input device.

Visually and hearing-impaired people must wear gloves, rings, or bracelets to input Finger Braille (5) into these devices. People who use these assistive devices (6) who are blind and deaf must not only wear the sensors, but also learn a new communication system. Children who have both hearing and vision impairments require different teaching strategies than children who only have one impairment. Although different countries define hearing and vision impairments differently, the negative effects of dual sensory impairments on access to environmental information are universally acknowledged, as is the fact that this unique disability necessitates specific teaching strategies to aid and support learning.

Because they cannot read text written on regular paper, visually impaired people use the Braille system in their writing and reading. However, Visual and hearing individuals in UAE are deprived of education and information due to a lack of basic reading materials, such as books and notes in Braille written format. As a result, we created an automatic Arabic speech to Braille translation system to address this issue by converting electronic Arabic books to Arabic Braille books.

Braille System

The Braille system (7) is a reading and writing system used worldwide by visually impaired people. Passing fingers across Braille characters or Braille cells is how it's read. In a Braille cell, six dots are arranged in a rectangle grid of two dots horizontally in a row and three dots vertically in a column, yielding four different patterns. Each dot arrangement is known as a cell (1), and each cell will have at least one elevated dot and a maximum of six.

Each of the twenty-five rows of text on a standard Braille printed page has forty cells. The size of a standard Braille page is about 11 inches by 11 inches. A Braille cell's size is also specified, though it varies greatly depending on the country. The dimensions of a Braille cell as printed on an embosser are shown in **Figure 1**.

The inventor of the Braille system anticipated the need to convey information in a compact format so that a group of cells could convey a lot more information in the string of letters that make up the cells. The cell's six dots (7) can be combined to make 63 different dot patterns. Despite the fact that there are a total of 64 patterns, the final one is a cell with no dots, which serves as a gap. When viewed through the lens of text representation in a computer, a Braille cell is thus analogous to a 6-bit character code. Almost all information can now be created electronically, and computer use in daily activities (typing and printing) has increased significantly. As a result, more specialized technologies that are accessible to visually impaired individuals are critical, allowing them to use computers for everyday transactions and contribute as much to society and the evolution of IT as others.

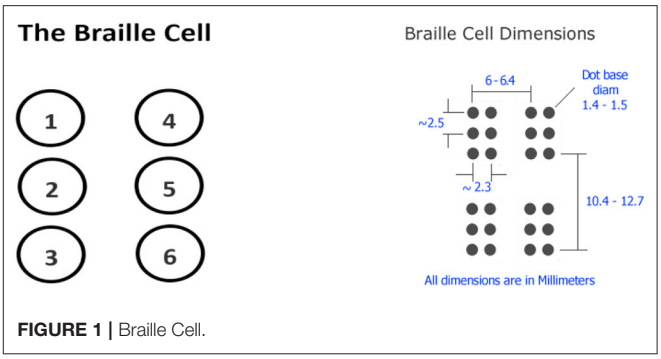


FIGURE 1 | Braille Cell.

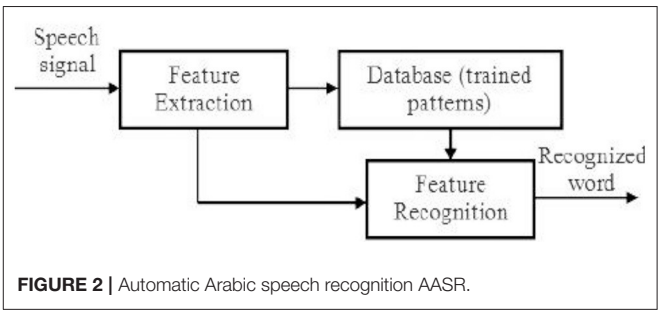


FIGURE 2 | Automatic Arabic speech recognition AASR.

Automatic Arabic Speech Recognition—AASR

Automatic Arabic Speech Recognition (AASR) (8) is a system that converts human speech into text. It's a difficult task because human speech signals are highly unpredictable due to a variety of speaker characteristics, speaking styles, and imprecise background noises (9). Variable-length voice signals must also be translated into variable-length word or label sequences by AASR. An AASR system is made up of many components, as shown in **Figure 2**. A microphone captures the speech signal first and foremost, which is then saved in digital format (10). The voice signal is non-stationary and fluctuates a great deal. As a result, the signal is divided into small chunks known as frames, with each frame's signal becoming more stationary and its features represented by a fixed length feature vector (11, 12). The procedure is referred to as feature extraction (13, 14), and the component that performs it is referred to as the front end. The front end's output is the observation sequence X, with each observation being a feature vector representing a single frame.

The acoustic model's goal is to simulate the acoustic characteristics of various speech units, as well as the relationship between the acoustic aspects of the entire voice utterance and the utterance label sequence or text. For spoken acoustic signals, the hidden Markov model (HMM) (15) is a popular and successful model. The HMM is very versatile when it comes to modeling variable length sequences, and it tolerates changes in speaking speeds.

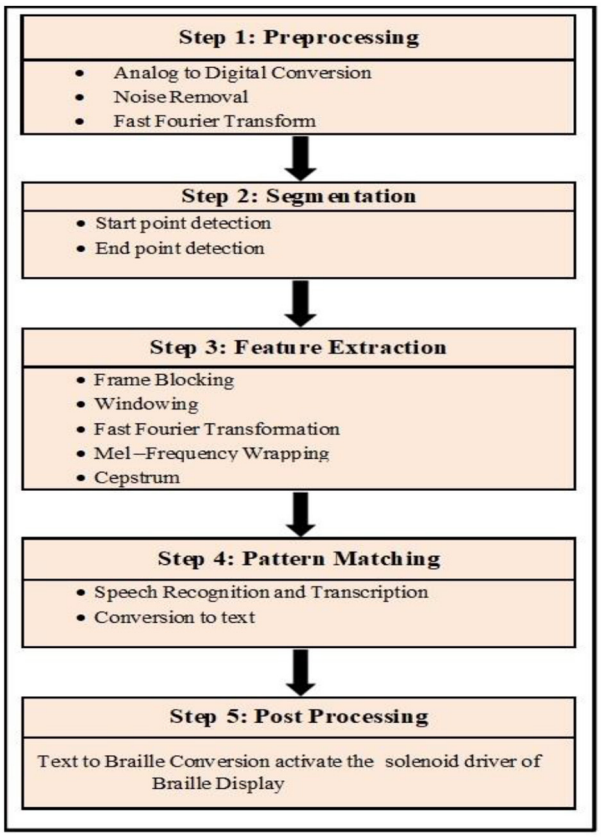


FIGURE 3 | Flow of Proposed Methodology.

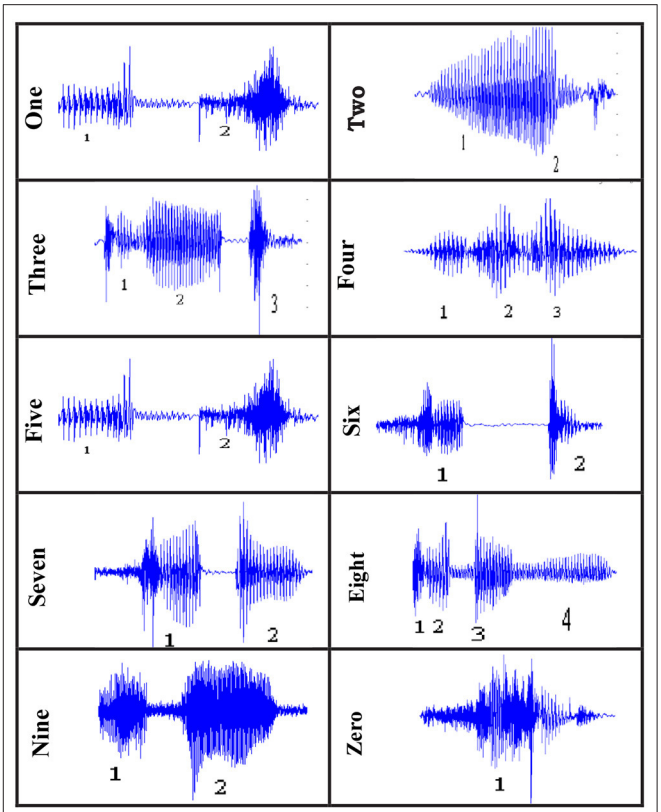


FIGURE 4 | Wave of Arabic Digit.

It models the likelihood of getting the speech observation sequence given a set of latent (hidden) discrete states. A decoder is used to recognize speech based on the acoustic and language models to model the relationship between an HMM state and the speech observations that pertain to this state. The decoder component searches for the best label sequence for a given voice utterance that maximizes the scores computed by the acoustic (16) and linguistic models. Because the number of viable label sequences grows exponentially with the length of the input feature sequence, the efficiency and speed of the decoder are critical elements of a successful ASR system (17).

Feature Extraction on ASR

The initial stage in speech recognition is the extraction (13, 14) of a sequence of feature vectors X that reflects the input voice signal. Feature vectors (18) are available in a wide range of sizes and shapes. In an acoustic model-based system, a good feature vector should include all relevant information for classification of the input signal (19). It's also a good idea to get rid of any data that isn't absolutely necessary. Mel frequency spectral coefficients (MFSC) and Mel-frequency cepstral coefficients (MFCC) are two feature vectors commonly used in academic and commercial ASR systems (20).

The first step is to take short overlapping windows of the speech signal, making it quasi-stationary and incorporating properties unique to single speech units. A Hamming window (or similar bell-shaped functions) is multiplied by this speech window to reduce the effect of discontinuity at the window's two edges. Because distinct speech units and sounds differ primarily in the distribution of energy along different frequencies, it is common practice to convert the speech time domain data into a frequency domain signal using Fast Fourier Transform (FFT) analysis (21). Filter bank analysis is then used to measure the energy in a small number of frequency ranges. These filters are dispersed along the Mel scale frequency to imitate human ear's sensitivity to frequency variances (22). Mel frequency filter bank analysis employs triangular filters. The energy in each filter is estimated to estimate the MFSC feature.

In practice, the log is used to calculate log-MFSC characteristics, which are closer to the sensitivity of the human ear and allow for better discrimination. In fact, neural network models, particularly convolutional neural networks, benefit from the log-MFSC feature. The log-MFSC features are closely correlated because the spectral envelop changes smoothly with frequency and neighboring filters overlap. The discrete cosine transform is used to minimize correlation in order to generate MFCC features (23). When using Gaussian Mixture Models (GMMs), this is especially important when using a diagonal covariance matrix.

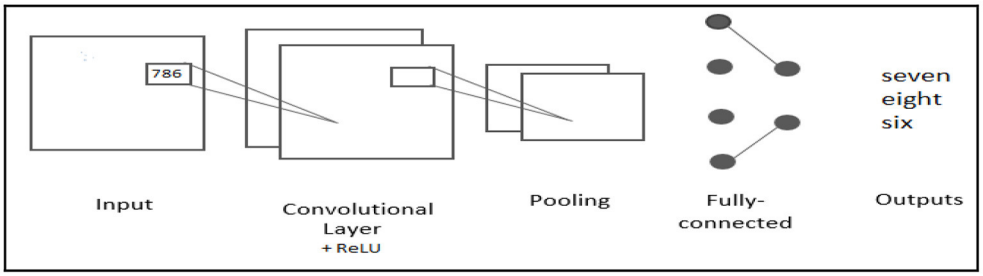


FIGURE 5 | CNN Architecture of Spoken Arabic digit to Text.

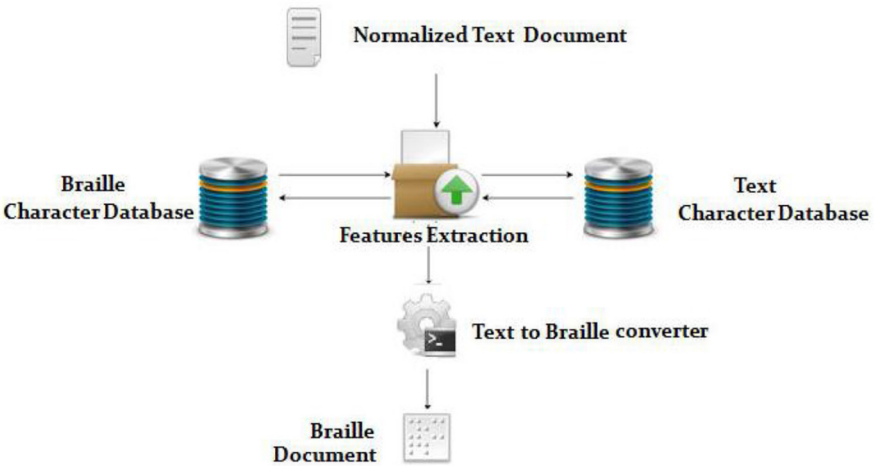


FIGURE 6 | Text to Braille Conversion.

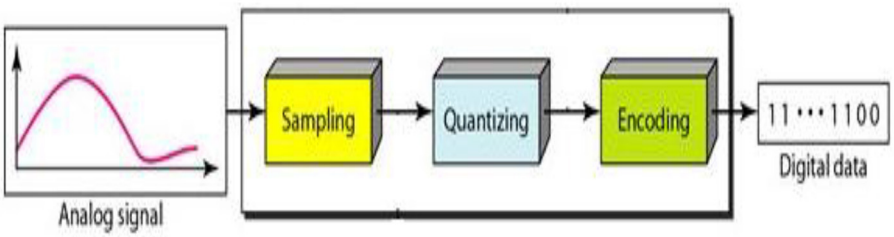


FIGURE 7 | Speech signal Capture.

HMM Based ASR System

An HMM can be used in a variety of ways to recognize speech. Isolated speech recognition is the most basic type, in which the system's goal is to classify individual words. In this scenario, each word is modeled using a different HMM model. The number of distinct acoustic portions of the word corresponds to the number of states in each HMM model. To classify the segment, the term associated with the model with the highest likelihood is used. The likelihood is calculated either exactly using the forward probability or roughly using the Viterbi algorithm's likelihood of the optimum state sequence in this scenario.

Speech recognition that is connected (continuous) is a more advanced technology that allows the speaker to continuously utter a large number of words. In this scenario, different word models are combined to create a large composite HMM. The types of transitions that are allowed between words are defined by a grammar. Furthermore, a language model is frequently used to assign different weights to different word transitions. Speech recognition is accomplished by determining the most likely state sequence and mapping that sequence into a word sequence that represents the uttered sentence. The disadvantage of this model is that learning a good model for each word necessitates a large

amount of training data. Furthermore, the model is incapable of recognizing terms not found in the training data.

To overcome these difficulties, models that represent sub-word units such as phonemes can be used. In these models, each phoneme is represented by a three-state sequence that includes the phoneme's transitional beginning and end, as well as the middle stationary state. To represent a word, a series of phoneme models are combined to form a word model. A dictionary is used to specify the phoneme models that represent a word. This method has the advantage of sharing each phoneme's training data across multiple words, reducing the amount of data needed for training. It can also recognize unseen words if they have equivalent entries in the dictionary.

Neural Network (NN) Based ASR

As a result of improved performance on a variety of speech recognition tasks, deep neural network models (DNNs) have recently resurfaced (26). Previous attempts to use NNs in acoustic modeling attempted to use the NN as a separate model for detecting simple speech units. To replicate the Viterbi algorithm's work in Lippmann (27), devised a recurrent neural architecture dubbed Viterbi net. As a result, it can recognize speech almost instantly. However, it lacked a training algorithm. The weights of the Recurrent Neural Network (RNN) were calculated using a trained HMM model. As a neural network architecture, the Time Delay Neural Network (TDNN) was proposed in Waibel et al. (28). The TDNN combines longer input contexts by using a series of consecutive input frames and hidden activations in the top layers. The model was used to classify a small number of phonemes. To successfully use NNs for ASR, a hybrid NN-HMM model was used. In one version of the model that has recently gained popularity, the NN replaces the GMM in scoring speech frames, while the HMM is used to model the speech signal's sequential temporal behavior.

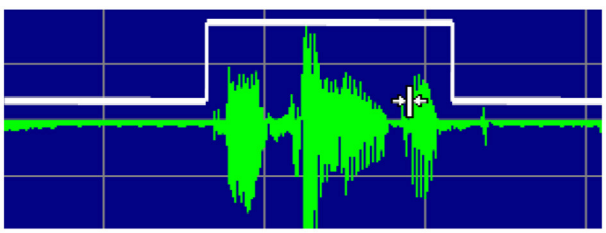


FIGURE 8 | Endpoint Identification.

RESEARCH METHODOLOGY

The Arabic language is the world's largest and most experienced Semitic language (26), and it differs from other European dialects such as English in several ways. Arabic is the Arab world's official language. Arabic is the world's sixth most communicated language, based on the number of people who speak it as their first language. Furthermore, ∼250 million people speak Arabic as their first language. Arabic sentences are written from right to left, and some letters may change shape depending on their position within a word (29).

Speech to-Braille STB (30) is a branch of computer science that focuses on developing computer frameworks that can recognize expressed words from an amplifier and convert them to messages. Speech to Text is the process of converting or deciphering an audible signal obtained by a receiver or a phone into a string

TABLE 1 | Performance Table.

Digit	Correct recognition	Wrong recognition	Accuracy of MFCC	Correct recognition (CNN and ReLU)	Accuracy After CNN and ReLU
1	15	5	75%	18	90%
2	13	7	65%	16	80%
3	13	7	65%	16	80%
4	15	5	75%	18	90%
5	13	7	65%	16	80%
6	12	8	60%	15	75%
7	14	6	70%	17	85%
8	13	7	65%	16	80%
9	14	6	70%	17	85%
0	16	4	80%	19	95%
Accuracy % (24)			69%	Accuracy % after CNN & ReLU	84%

TABLE 2 | Comparison of Accuracy.

Models	Dataset	Features	Recognizer	Accuracy
Model in Rajagede et al. (25)	Ten Arabic letters	MFSC+delta	CNN	80.75%
Model in Wazir and Chuah (24)	Arabic Digit 0–9	MFCC	LSTM	69%
Proposed Model	Arabic Digit 0–9	MFCC	CNN +ReLU	84%

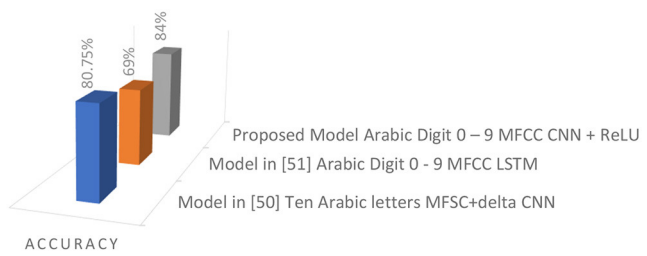


FIGURE 9 | Comparison of Accuracy.

CHARACTER	BRAILLE	BRAILLE DOTS
number sign		3456
١ 1		3456 1
٢ 2		3456 12
٣ 3		3456 14
٤ 4		3456 145
٥ 5		3456 15
٦ 6		3456 124
٧ 7		3456 1245
٨ 8		3456 125
٩ 9		3456 24
٠ 0		3456 245

FIGURE 10 | Arabic Digit to Braille.

of words. Extraction of highlights from a gathered signal and matching them to a voice test in an information base is used to perceive a word (31, 32).

The goal of the procedure is to create a Digit speech recognition refers to the recognition of utterances in which the user can speak normally without pausing between words, which is similar to how humans speak. Subword sounds, such as phonemes (alphabets in Indian languages) (29), are commonly used as the basic unit of recognition in continuous speech recognition systems. The sequence of phonetic models forms the larger linguistic unit, such as words/sentences, during recognition. Because the number of sub-word units is limited, collecting enough data to train each of these basic sounds in a variety of acoustic environments is not difficult. Adding a new word is simple because all that is required is a description of the new word in terms of previously learned speech sounds. Among the features are speech analysis, training tools, recognition tools, outcome analysis, and labeling tools. It can recognize both single and related words. It is capable of modeling both full words and sub-word units.

The system's post-processing involves retrieving the exact data associated with the recognized speech. As shown in **Figure 3**, the fetched Arabic data corresponding to the Arabic Corpus dataset is converted into Arabic Braille script and delivered to the user in the form of Braille Display.

CONVOLUTIONAL NEURAL NETWORKS FOR ASR

Speech signals can take many different forms that have little to do with the speech's grammatical content. For example, different speakers may have different voice qualities. There can be a lot of differences between speakers, even if they are the same person. Variations in speaking speed, pitch, and formant frequencies due to changes in vocal tract length, as well as background noise, are examples.

A variety of strategies have been proposed to deal with these differences. In this case, there are two approaches: passive and active. The goal of passive strategies is to standardize speech variances. This can be accomplished by employing special features such as MFCCs, which filter out information that is unrelated to the speech's linguistic content. The HMM normalizes rate variations because it is tolerant of variations in speaking pace.

Active approaches aim to estimate some parameters that transform the speech signal or to fine-tune the trained speech models to perform better in new situations. Some speaker adaptation strategies, for example, change the HMM model parameters to reflect the new speaker's characteristics. Other speaker adaptation strategies convert a new speaker's speech signal into a canonical speaker that the acoustic model can learn from. The new model or transformation parameters are calculated from some speech adaption data collected from the new speaker in both situations before performing speech recognition on this speaker.

The Convolutional Neural Network (33) (CNN) is a type of artificial neural network that processes data in the form of patterns. CNN is an effective method for building deep learning models that have been used in object recognition applications. CNN (34) was proposed as a method of reducing the amount of data required for traditional artificial neural networks. Three layers make up a traditional Convolutional Neural Network: convolution, pooling, and fully connected layers (35). The pooling layer is used to reduce the number of values in feature maps, while the convolution layer is used to extract features. The best feature maps are passed into the classifier using a fully linked layer (35, 36). CNN is scalable in general and the training procedure takes less time.

The main advantage of the CNN is that domain knowledge can be used to create the CNN structure in order to manage

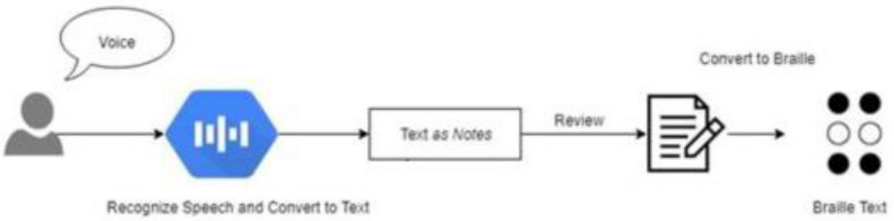


FIGURE 11 | Spoken Arabic Digit to Braille.

application-specific variances. In these applications, CNNs are used to take advantage of translation invariance and to tolerate small changes in visual patterns in multiple directions. Weight sharing, local connectedness, and pooling are some of the network topologies used by CNNs, as opposed to standard NNs. As a result, minor image deformations and noise, as well as small rotations and scaling, are less noticeable to the CNN. These invariances are difficult to learn automatically in typical NNs. In a similar way, CNN ideas have recently been applied to voice processing.

Unsupervised learning of acoustic features was achieved using the CNN structure (37). Convolution over time is used in these studies to manage small temporal shifts for better speaker, gender, and phonetic classification representations, but not for audio recognition.

The CNN structure allows us to acquire frequency invariant features when applied along the frequency axis (38, 39). A CNN can also accommodate temporal changes and thus better handle speaking speed variances when applied along the time axis (40). However, the HMM can handle these variances inside the hybrid NN-HMM model, reducing the benefit of applying convolution and pooling over time.

Description of Arabic Digits

Due to background noise interference, speech is harmed in voice applications. As a result, we can't tell whether a signal contains valid data just by looking at it (41). Many efforts have been made in speaker system independent, automatic Arabic speech recognition for an isolated digit or continuous digit speech for many languages. There is still a scarcity of research on speech recognition in the Arabic language. All speech recognition applications, including those described here, are, however, available in English. Despite the fact that our study focused on Arabic speech recognition in digit contexts (**Figure 4**).

Feature Extraction and Classification Using CNN

The feature extraction technique for audio data is depicted in **Figure 5** (42–46). It shows the spectrum of audio data before and after feature extraction using the Python'sklearn' module. The data was split 80:20, with 80% going into the training dataset and 20% into the test dataset, with 20% of the training dataset being used as a validation set for this Speech recognition processing (47). The model processes the training, test, and validation datasets, which will be used in the trained model's test dataset prediction analysis.

- In the convolution layers, we created two separate models, each with its own set of parameters.
- In both models, the first 2D-convolutional layer (conv2D) was used to analyze train data, and it has a 128-bit resolution (Model-1).
- A RELU activation function with a filter size of 64 (model-2).
- The output of the first Conv2D layer was sent to the first batch normalization layer.
- For each model, this procedure is repeated twice using the second and third conv2D.
- The first model employs 48 filters with a normalization pool size of (3, 3), whereas the second employs 32 filters with the remaining two parameters unchanged.

After the feature learning stage, the classification technique starts by flattening the resulting vector into a one-dimensional vector. The data was flattened because categorization is easier with a one-dimensional vector. The 1D vector of the flatten layer is then passed through a series of layers.

Finally, the output vector is sent to a 20-unit dense layer, which corresponds to the number of key words in the Arabic dataset. After that, the final result was obtained. The optimizer in this model is 'Adam,' and the loss function is 'Sparse-Categorical Cross-entropy.' **Figure 5** depicts the basic architecture of our proposed model.

TEXT TO BRAILLE

Figure 6 depicts the steps involved in converting text to braille.

- Enter Arabic Text: Entering Arabic text in the provided area or by browsing and attaching a text file (48).
- Validation and data verification of Arabic text: The primary function of this block is to validate and verify the Arabic content extracted from the speech. Validating a meaningful text and verifying the converted word in a database (Text Database) (48) is part of the validation procedure.
- Braille conversion process: The Braille Process converts each character into an equivalent Braille character. The symbols for each and every Arabic digit character can be found in the currently available database (Braille Database) (49–51).

- Arabic Braille Output: After the conversion has been replaced appropriately, the final output, as shown in **Figure 6**, can be handled through the output interface to reach the end user.

EXPERIMENTAL ANALYSIS

The Methodology of Speech recognition and Conversion to Braille system (30) has following steps.

Step 1: The first step in this research is to gather information from a variety of Arabic speakers. Speakers must record the Arabic numerals 0 to 9 and say each Arabic number separately in a clear voice with minimal background noise, as shown in **Figure 7**. As a result, data collection was required in order to develop the speech recognition system; this data was collected by various countries around the world, each with its own dialect for pronouncing words (24, 25). Geographical areas influence how speech is delivered, but they aren't always the source of variation; for example, dialects may be influenced by family history. The voice signal is obtained by recording or capturing the entire audible frequency range of a human voice. The sampling frequency is defined as the number of samples obtained in a second.

Step 2: To define where a single utterance begins and ends, the zero-crossing rate (ZCR) and average amplitude thresholds are used. It can be classified as a speech signal segment if the energy in the frame exceeds the energy threshold. However, because unvoiced speech signals are sometimes merged or combined with ambient noises, the marking of start and stop locations may not be precise. Unvoiced speech has a higher ZCR than noise, so it can be used to distinguish between noise and a weak speech signal, allowing for better start and stop point marking, as shown in **Figure 8**.

Step 3: Using CNN to recognize speech: The classification stage is followed by a fully connected layer and a softmax layer, and the convolutional layers and pooling layers are used to learn features. Except for the fully-connected layer, all convolutional and pooling layers used the ReLU activation function. To extract features, have used the following settings: 16,000 sample rate, 512 and 160 sample FFT lengths and hop sizes, as well as the Hamming window function The first CNN layer conducts a convolutional over the input feature vector using 32 ReLU kernels with 3×3 receptive fields and a stride of 1 in both dimensions, as shown in **Table 1**.

The proposed model was tested using an Arabic speech corpus of isolated words. The used corpus is subjected to a variety of modifications, including pitch, speed, dynamic range, noise, and time shifting forward and backward. As shown in **Table 2** and **Figure 9**, the maximum accuracy achieved when using the ReLU activation function with CNN was 84%. The ReLU activation function, a non-linear activation function, has gained prominence in this proposed methodology. When comparing the results of this study to previous research, it is clear that the proposed methodology performed better in the AASR.

Step 4: Text to Braille Conversion: The result, as shown in **Figure 10**, will be processed through the Arabic Text database, which will identify the Arabic text.

Step 5: Postprocessing: The Arabic letter is embossed as Arabic Braille script on the Braille display using solenoid drive (52) as shown in **Figure 11**. The evaluation is based on how it is handled and prioritized including some other performance parameters. The primary reason for employing the device would be decided by determining if the device's benefit outweighs the negative encountered. Following the prototype's implementation, it is evaluated with a small group of visually challenged people, as well as people who are both visually and hearing impaired.

CONCLUSIONS

Braille is a text-based communication system that is solely dependent on touch. People with vision and hearing impairments would use it. We propose a Speech to Braille conversion procedure to provide the visually and hearing-impaired people can easily make notes in Braille using the provided methods (53) by simply speaking or playing a tape. This app will help people with both visual and hearing impairments communicate in real time, and it will be a cost-effective way. This research produces output that is comparable to traditional high-cost braille printers, as well as assisting visually and hearing-impaired people in recognizing the Speech and converting into Braille output. Speech recognition is a bonus feature that allows visually and hearing-impaired people to print braille for personal use without a computer or a mediator by simply speaking the phrases to be printed into a microphone.

Arabic digits six are particularly difficult to recognize when spoken. The complexity of the voiced signals of the spoken digits is the reason. Deep learning requires a large amount of data to achieve better accuracy, the accuracy of CNN model could be greatly improved by using a large amount of training data. More digit voice samples with noise, for example, could give us a more robust classifier. It's worth noting that as the amount of training data grows, so does the amount of time spent training.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AD and SB: conceptualization, investigation, and methodology. SB and RA: funding acquisition and resources. AM: project administration and visualization. AD: writing of the original draft. AM and SB: writing of the review and editing. SB: validation. RA: data curation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Project No. AN000602).

REFERENCES

1. Fukushima S. *Person With Deafblind and Normalization*. Tokyo: Akashi Shoten (1997).
2. Uehara N, Aoki M, Nagashima Y, Miyoshi K. Bidirectional telecommunication using wearable i/o terminal for finger braille. In: *Proceedings of the Human Interface Symposium* (San Diego, CA). (2000). pp. 37–40.
3. Hoshino T, Otake T, Yonezawa Y. A study on a finger-braille input system based on acceleration of finger movements. In: *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol. 85*. (2002). p. 731.
4. Ochi T, Kozuki T, Suga H. Bracelet type braille interface. *Correspond Hum Interf.* (2003) 51:25–7. doi: 10.5772/51472
5. Matsuda Y, Isomura T, Sakuma I, Kobayashi E, Jimbo Y, Arafune T. Finger Braille teaching system for people who communicate with deafblind people. In: *2007 International Conference on Mechatronics and Automation*. Heilongjiang: IEEE (2007). pp. 3202–7.
6. Matsuda Y, Sakuma I, Jimbo Y, Kobayashi E, Arafune T, Isomura T. Development of finger Braille teaching system-teaching of dotting finger and position using speech recognition. *J Life Support Eng.* (2007) 19:105–16. doi: 10.1109/ICMA.2007.4304074
7. Baboo CD, Devi VA. Computer based Tamil Braille system–a review. In: *11th International Tamil Internet Conference, Annamalai University, Chidambaram, India, Jointly Conducted by International Forum for Information Technology (INFITT) and Annamalai University* (Tamil Nadu). (2012). pp. 57–61)
8. Khilari P, Bhope VP. A review on speech to text conversion methods. *Int J Adv Res Comput Eng Technol.* (2015) 4:3067–72. Available online at: https://www.irjet.net/archives/V7/i5/IRJET-V7I5854.pdf
9. Devi VA, Suganya MV. An analysis on types of speech recognition and algorithms. *Int J Comp Sci Trends Technol.* (2016) 4:350–5. Available online at:http://www.ijcstjournal.org/volume-4/issue-2/IJCST-V4I2P62.pdf
10. Ping Z, Li-Zhen T, Dong-Feng X. Speech recognition algorithm of parallel subband HMM based on wavelet analysis and neural network. *Inform Technol J.* (2009) 8:796–800. doi: 10.3923/itj.2009.796.800
11. O'Shaughnessy D. Interacting with computers by voice: automatic speech recognition and synthesis. *Proc IEEE.* (2003) 91:1272–305. doi: 10.1109/JPROC.2003.817117
12. Anusuya MA, Katti SK. Front end analysis of speech recognition: a review. *Int J Speech Technol.* (2011) 14:99–145. doi: 10.1007/s10772-010-9088-7
13. Jung S, Son J, Bae K. Feature extraction based on wavelet domain hidden Markov tree model for robust speech recognition. In: *Australasian Joint Conference on Artificial Intelligence.* Berlin; Heidelberg: Springer (2004). pp. 1154–9.
14. Sárosi G, Mozsáry M, Mihajlik P, Fegyó T. Comparison of feature extraction methods for speech recognition in noise-free and in traffic noise environment. In: *2011 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD).* Brasor: IEEE (2011). pp. 1–8.
15. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE.* (1989) 77:257–86. doi: 10.1109/5.18626
16. Walker W, Lamere P, Kwok P, Raj B, Singh R, Gouvea E, et al. Sphinx-4: A flexible open source framework for speech recognition.
17. Almeman K, Lee M, Almiman AA. Multi dialect Arabic speech parallel corpora. In: *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA).* Sharjah: IEEE (2013). pp. 1–6.
18. Amin A, Mari JF. Machine recognition and correction of printed Arabic text. *IEEE Transac Syst Man Cybernet.* (1989) 19:1300–6. doi: 10.1109/21.44052
19. Sharma HV, Hasegawa-Johnson M. Acoustic model adaptation using in-domain background models for dysarthric speech recognition. *Comp Speech Lang.* (2013) 27:1147–62. doi: 10.1016/j.csl.2012.10.002
20. Awaidah SM, Mahmoud SA. A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models. *Signal Process.* (2009) 89:1176–84. doi: 10.1016/j.sigpro.2008.12.022
21. Tiwari V. MFCC and its applications in speaker recognition. *Int J Emerg Technol.* (2010) 1:19–22. Available online at: https://www.semanticscholar.org/paper/Finger-Braille-Teaching-System-173-Finger-Braille-Matsuda-Isomura/8192b9983f231fcb6bf971cf860b8cebb5a3f3f0
22. Farooq O, Datta S. Mel filter-like admissible wavelet packet structure for speech recognition. *IEEE Signal Proc Lett.* (2001) 8:196–8. doi: 10.1109/97.928676
23. Chen C, Bunescu RC, Xu L, Liu C. Tone classification in Mandarin Chinese using convolutional neural networks. In: *Interspeech* (San Francisco, CA). (2016). pp. 2150–4.
24. Wazir AS, Chuah JH. Spoken arabic digits recognition using deep learning. In: *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS).* Selangor: IEEE (2019). pp. 339–44.
25. Rajagede RA, Dewa CK. Recognizing Arabic letter utterance using convolutional neural network. In: *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD).* Kanazawa: IEEE (2017). pp. 181–6.
26. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Proc Mag.* (2012) 29:82–97. doi: 10.1109/MSP.2012.2205597
27. Lippmann RP. Neural-net classifiers useful for speech recognition. In: *Proceedings International Conference on Neural Networks, Vol. 4* (Piscataway, NJ). (1987). pp. 417–25.
28. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ. Phoneme recognition using time-delay neural networks. *IEEE Transac Acoustics Speech Signal Process.* (1989) 37:328–39. doi: 10.1109/29.21701
29. Elmahdy M, Gruhn R, Minker W. *Novel Techniques for Dialectal Arabic Speech Recognition*. Springer Science & Business Media (2012).
30. Devi VA. Conversion of speech to Braille: interaction device for visual and hearing impaired. In: *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN).* Tamil Nadu: IEEE (2017). pp. 1–6).
31. Ajami Alotaibi Y, Alghamdi M, Alotaiby F. Speech recognition system of Arabic alphabet based on a telephony Arabic corpus. In: *International Conference on Image and Signal Processing.* Berlin; Heidelberg: Springer (2010). pp. 122–9.
32. Alotaibi YA. Investigating spoken Arabic digits in speech recognition setting. *Inform Sci.* (2005) 173:115–39. doi: 10.1016/j.ins.2004.07.008
33. Abdel-Hamid O, Mohamed AR, Jiang H, Deng L, Penn G, Yu D. Convolutional neural networks for speech recognition. *IEEE ACM Trans Audio Speech Lang Process.* (2014) 22:1533–45. doi: 10.1109/TASLP.2014.2339736
34. Nazih W, Hifny Y, S Elkilani W, Mostafa T. Fast detection of distributed denial of service attacks in VoIP networks using convolutional neural networks. *Int J Intellig Comput Inform Sci.* (2020) 20:125–38. doi: 10.21608/ijicis.2021.51555.1046
35. Yamashita R, Nishio M, Do RK, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging.* (2018) 9:611–29. doi: 10.1007/s13244-018-0639-9
36. Abeßer J. A review of deep learning based methods for acoustic scene classification. *Applied Sciences.* (2020) 10:2020. doi: 10.3390/app10062020
37. Lee H, Pham P, Largman Y, Ng A. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Adv Neural Inform Process Syst.* (2009) 22:1–5. doi: 10.1109/ICASSP.2013.6638952
38. Abdel-Hamid O, Mohamed AR, Jiang H, Penn G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Kyoto: IEEE (2012). pp. 4277–80.
39. Deng L, Abdel-Hamid O, Yu D. A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing.* Vancouver, BC: IEEE (2013). pp. 6669–73.
40. Abdel-Hamid O, Deng L, Yu D. Exploring convolutional neural network structures and optimization techniques for speech recognition. In: *Interspeech, Vol. 2013* Lyon. (2013). pp. 1173–5.
41. Veena KV, Mathew D. Speaker identification and verification of noisy speech using multitaper MFCC and Gaussian Mixture models. In: *2015 International Conference on Power, Instrumentation, Control and Computing (PICC).* Thrissur: IEEE (2015). pp. 1–4.
42. Hammami N, Bedda M. Improved tree model for Arabic speech recognition. In: *2010 3rd International Conference on Computer*

Science and Information Technology. Vol. 5. IEEE (2010). pp. 521–6.
43. Mahmood A, Alsulaiman M, Muhammad G. Automatic speaker recognition using multi-directional local features (MDLF). *Arabian J Sci Eng.* (2014) 39:3799–811. doi: 10.1007/s13369-014-1048-0
44. Elkourd A. Arabic isolated word speaker dependent recognition system. In: *Source Framework for Speech Recognition*. (2004).
45. Alsulaiman M, Mahmood A, Muhammad G. Speaker recognition based on Arabic phonemes. *Speech Commun.* (2017) 86:42–51. doi: 10.1016/j.specom.2016.11.004
46. Khelifa MO, Elhadj YM, Abdellah Y, Belkasmi M. Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system. *Int J Speech Technol.* (2017) 20:937–49. doi: 10.1007/s10772-017-9456-7
47. Bhalla K, Koundal D, Bhatia S, Khalid M, Rahmani I, Tahir M. Fusion of infrared and visible images using fuzzy based siamese convolutional network. *Comput. Mater. Con.* (2022) 70:5503–18. doi: 10.32604/cmc.2022.021125
48. Cong L, Asghar S. Robust speech recognition using fuzzy matrix/vector quantisation, neural networks and hidden Markov models. In: *International Conference on Advances in Pattern Recognition.* London: Springer (1999). pp. 255–64.
49. Merazka F. VQ Codebook design using genetic algorithms for speech line spectral frequencies. In: *International Symposium on Intelligence Computation and Applications.* Berlin; Heidelberg: Springer (2012). pp. 557–66.
50. Samal A, Parida D, Satapathy MR, Mohanty MN. On the use of MFCC feature vector clustering for efficient text dependent speaker recognition. In: *Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA).* Cham: Springer (2013) pp. 305–12.
51. Ananthi S, Dhanalakshmi P. SVM and HMM modeling techniques for speech recognition using LPCC and MFCC features. In: *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA).* Cham: Springer (2014). pp. 519–26.
52. Baboo CD, Devi VA. Tamil braille system: A conversion methodology of tamil into contracted braille script (grade2). In: *UACEE International Journal of Advances in Computer Science and its Applications - IJCSIA, Vol. 3.* (2013).
53. Devi A. *Portable Device Converting Braille and Text Document to Speech Output Using Raspberry Pi.* IJCTA (2016). pp. 4809–18.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bhatia, Devi, Alsuwailem and Mashat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fpubh.2022.879418

Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography

Fatemeh Sharifonnasabi[1], Noor Zaman Jhanjhi[1], Jacob John[2], Peyman Obeidy[3], Shahab S. Band[4]*, Hamid Alinejad-Rokny[5,6,7] and Mohammed Baz[8]*

[1] Department of Computer Science & Engineering, School of Computing & IT (SoCIT), Taylor's University, Subang Jaya, Malaysia, [2] Department of Restorative Dentistry, Faculty of Dentistry, University of Malaya, Kuala Lumpur, Malaysia, [3] Charles Perkins Centre, Faculty of Medicine and Health, University of Sydney, Darlington, NSW, Australia, [4] Future Technology Research Centre, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan, [5] BioMedical Machine Learning Lab (BML), The Graduate School of Biomedical Engineering, University of New South Wales (UNSW) Sydney, Kensington, NSW, Australia, [6] UNSW Data Science Hub, The University of New South Wales, UNSW Sydney, Kensington, NSW, Australia, [7] Health Data Analytics Program, AI-enabled Processes (AIP) Research Centre, Macquarie University, Macquarie Park, NSW, Australia, [8] Department of Computer Engineering, College of Computer and Information Technology, Taif University, Taif, Saudi Arabia

Age estimation in dental radiographs Orthopantomography (OPG) is a medical imaging technique that physicians and pathologists utilize for disease identification and legal matters. For example, for estimating post-mortem interval, detecting child abuse, drug trafficking, and identifying an unknown body. Recent development in automated image processing models improved the age estimation's limited precision to an approximate range of +/- 1 year. While this estimation is often accepted as accurate measurement, age estimation should be as precise as possible in most serious matters, such as homicide. Current age estimation techniques are highly dependent on manual and time-consuming image processing. Age estimation is often a time-sensitive matter in which the image processing time is vital. Recent development in Machine learning-based data processing methods has decreased the imaging time processing; however, the accuracy of these techniques remains to be further improved. We proposed an ensemble method of image classifiers to enhance the accuracy of age estimation using OPGs from 1 year to a couple of months (1-3-6). This hybrid model is based on convolutional neural networks (CNN) and K nearest neighbors (KNN). The hybrid (HCNN-KNN) model was used to investigate 1,922 panoramic dental radiographs of patients aged 15 to 23. These OPGs were obtained from the various teaching institutes and private dental clinics in Malaysia. To minimize the chance of overfitting in our model, we used the principal component analysis (PCA) algorithm and eliminated the features with high correlation. To further enhance the performance of our hybrid model, we performed systematic image pre-processing. We applied a series of classifications to train our model. We have successfully demonstrated that combining these innovative approaches has improved the classification and segmentation and thus the age-estimation outcome of the model. Our findings suggest that our innovative model, for the first time, to the best of our knowledge, successfully estimated the age in classified studies of 1 year old, 6 months, 3 months and 1-month-old cases with accuracies of 99.98, 99.96, 99.87, and 98.78 respectively.

Keywords: dental age, estimation, Orthopantomogram, convolutional neural network, k-nearest neighbor, Biomedical machine learning

OPEN ACCESS

Edited by:
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Ayse Oktay,
Yıldız Technical University, Turkey
Dirk Vandermeulen,
University Hospitals Leuven, Belgium
Mohd Yusmiaidil Putera Mohd Yusof,
Universiti Teknologi MARA, Malaysia
Ehsan Kozegar,
University of Guilan, Iran

***Correspondence:**
Shahab S. Band
shamshirbands@yuntech.edu.tw
Noor Zaman Jhanjhi
NoorZaman.Jhanjhi@taylors.edu.my

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 February 2022*
Accepted: *22 April 2022*
Published: *30 May 2022*

Citation:
Sharifonnasabi F, Jhanjhi NZ, John J, Obeidy P, Band SS, Alinejad-Rokny H and Baz M (2022) Hybrid HCNN-KNN Model Enhances Age Estimation Accuracy in Orthopantomography. Front. Public Health 10:879418. doi: 10.3389/fpubh.2022.879418

INTRODUCTION

Age estimation in living and deceased individuals has always been important in pediatric studies, pathological complications, and forensic medicine (1, 2). Human bone age can be estimated from foot, shoulder, ankle, hip, elbow, cervix, ankle, and teeth bones (3–5). The accuracy of the calculated age from a bone specimen depends on genetic and environmental factors such as age, race, smoking, lifestyle, famine and natural disasters (6). While current age estimation methods, for example, Greulich-Pyle or Tanner–Whitehouse method, can accurately estimate age from radiological examination of the left hand in children, age assassination complexities significantly increase when comparing young adults to adolescents. The absence of formal age documentation makes the process even more complicated despite recent discoveries in anthropometric fields demonstrating the importance of other bones, such as the pelvis, in determining bone age (3, 4). OPG or dental radiographs remained to be the primary forensic concern (7).

Orthopantomography, also known as an OPG or dental radiographs, provides a wide panoramic view of the lower face bones (8). OPG displays all the teeth on both jaws. OPG also depicts the jawbone and the temporomandibular joint (TMJ) on the same radiography film. The dental OPG images have been suggested to be more effective when compared to other bones of the body X-ray images. Teeth bones are long-lasting and resistant to high temperatures (2) and organisms which decompose a dead body (9). The Malaysian Institute of Forensic Medicine claimed that the X-ray images of teeth are the primary and the most used method in estimating age (10, 11). However, the traditional image processing approach for OPG age estimation is still relatively expensive, laborious and requires long-term monitoring, increasing potential radiation exposure (12).

Traditional image processing for dental age measurement is a manual method that may include several steps such as segmentation, feature extraction, image pre-processing, classification, or regression. Each of these steps is error-prone and can induce variations in the outcome. For example, in an identical age range, the dry bone image will be different from the wet bone in radiography scan. In recent years, Machine Learning/Deep Learning techniques have been widely used to identify patterns in complex data such as clinical imaging, genomics, bioimaging, and phenotypic data (13–17). Machine learning algorithms for example, NN, SVM, KNN, Decision Tree, displayed promising ability in prediction and classification (18–28) including estimation of bone age (29, 30). The medical and biological datasets are increasing rapidly. To analyse such big and complex data, artificial intelligence and machine learning algorithms become most popular (31–41). Therefore, it is important to implement novel techniques to uncover the medical and biological patterns. In particular machine learning and deep learning techniques have been widely used to analyse imaging data (23, 37, 42).

Deep learning-based methods, also known as end-to-end learning-based methods, such as convolutional neural networks, are either unsupervised that operate directly on the input images and generate the desired output without intermediate steps such as segmentation and extraction or supervised. Supervised learning (SL) models utilize a learning function that requires example input-output pairs for the model validation. The supervised approach runs the comparison between a radiographic image of each subject to an existing reference (i.e., labeled images), including gender and age information (43). Nevertheless, most of these methods require a large amount of labeled data.

Collecting and labeling large datasets are usually expensive, time-consuming, or sometimes impossible. Thus, in recent years, Semi-Supervised Learning (SSL) has emerged. Leveraging both labeled and unlabelled data, SSL has been proved to be a practical approach. Transfer learning (TL), an SSL-based model) aiming to utilize training knowledge gained from one data set to analyse another using a few-shot classification framework (44, 45). Moreover, unprocessed image databases, which usually can be fed to conventional ML or DL models, are prone to inaccurate class recognition. Extracting specific characteristics that accurately represent behavior in different environmental conditions is also a time-consuming task. (11, 46). Hence, choosing a classifier that can differentiate behaviors in terms of diversity within and outside each class is one of the challenges in recognizing behavior. Therefore, converting raw data into attribute vectors systematically and using effective attribute extractors using an engineering approach is required for proper classification (47–49).

Despite these improvements developing and training deep neural networks remained challenging and time-consuming. Thus, Transfer learning (TL) can use the pre-trained deep network to perform data classification, attracting increasing attention (50, 51) this article proposed an innovative automated machine learning model approach to estimate bone age. This approach is based on a deep convolutional neural network (52). We improved the accuracy of age determination in dental panoramic images with approximately 6-month intervals. The proposed method determines the precise age in the range of 15 to 23 years old that is divided into nine age groups. Each age group includes subcategories of images with 1 year, six, three and 1-month(s) intervals. This method combined the methodology of features extraction of the convolutional neural networks with the nearest neighbor feature to analyse and classify information and features in dental images.

More specifically, the architecture of the proposed HCNN-KNN model used in this study consists of four steps (**Figure 1**). Initially, a convolutional neural network (CNN) model was trained on OPG images. Then, a fully connected layer generated from CNN was fed into a principal component analysis (PCA). The PCA performed data transformation and dimension reduction. In the last step, data were classified using the KNN algorithm. The details of the prediction process of the classes in the data set were explained in the following sections.

Related Literature

Age estimation based on dental X-ray images remained a critical field to develop. Tooth bones are immortal and can last more than 20 years through post-mortem decomposition. This made the tooth an essential part of the forensic investigation

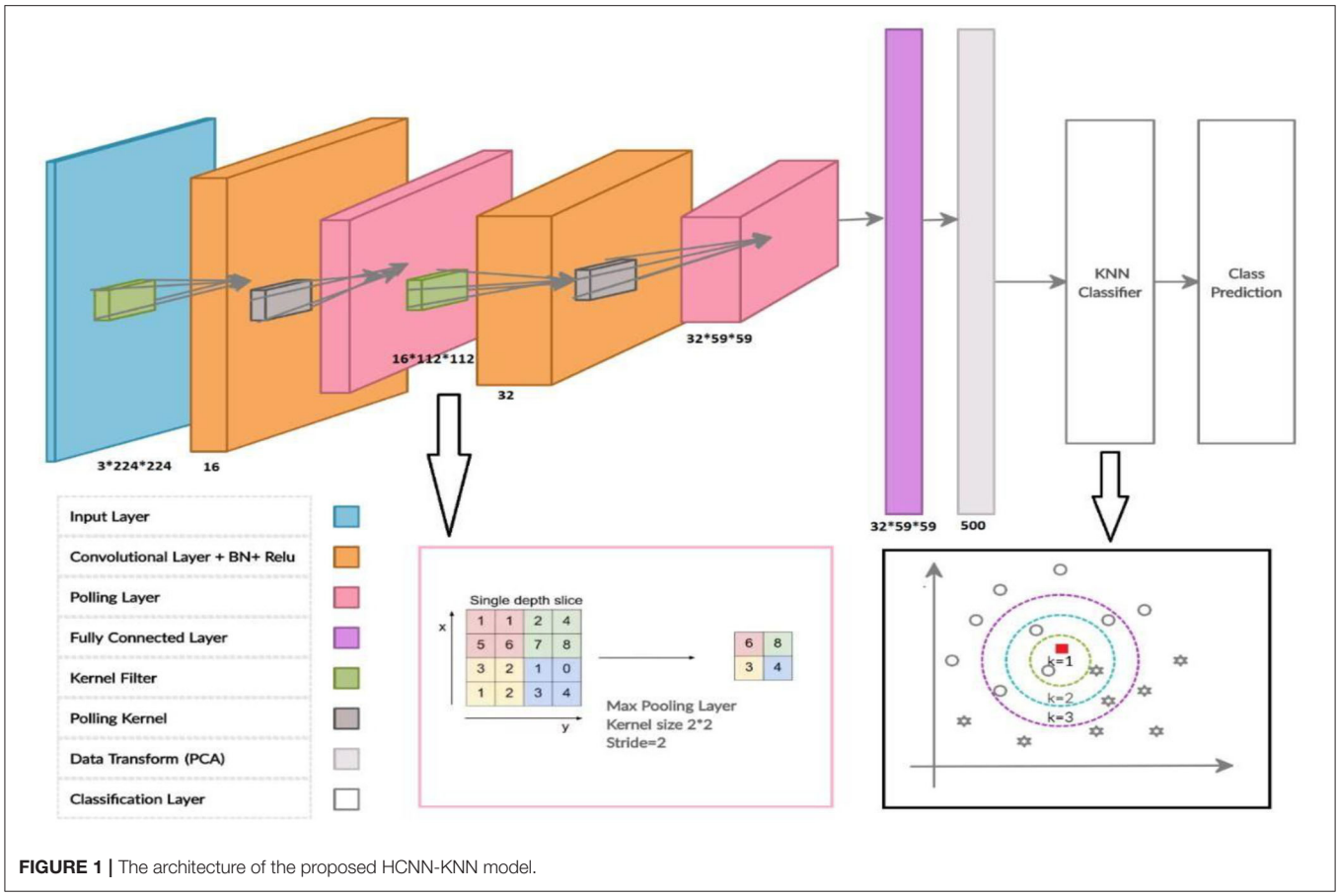


FIGURE 1 | The architecture of the proposed HCNN-KNN model.

(53). Although the number of bone age estimation models has increased, most remain in the research state with slow or no transition in the industry. Computerized methods for bone age estimation based on OPG images are currently limited to ± 6 months precision (52). Moreover, current Deep neural network algorithms used to estimate the age using dental X-ray images are, for example, AlexNet and ResNet. Training these algorithms requires large imaging datasets. For instance, Houssein et al. used a dataset with 1,429 dental X-ray images to reach acceptable output parameters (11).

In another study, Avuçlu et al. (54) used 1,315 dental images and 162 different dental classes to morphologically predict the age and gender in dental X-ray images with 95% accuracy (54, 55). Farhadian et al. (30) applied Transfer learning to reduce using a neural network that relied on dental data to alter age assessment. The age group of between 14 and 60 was used, and the overall data sample size for the research was 300 images. With an MAE of 4.12 years, the neural network methodology showed a reasonably accurate outcome (30). The neural network errors were smaller than the regression model's estimation errors, with the test data set being RMSE of 10.26 years and MAE of 8.17 years.

Alkaabi et al. (8) examined different Convolutional Neural Networks s AlexNet, VGGNet, and ResNet. Architectures for age estimation. They used common CNN architectures to perform forensic dentistry's automatic age calculation without any modifications. Using Capsule-Net, predicting age estimation from dental images was also performed, which depicted 36% higher accuracy than the CNNs model. The Capsule-Net model based on transfer learning reached a cumulative accuracy of 76% for ± 1-Year-old sample OPG images.

Tao et al. (29) has introduced the Multilayer Perceptron Neural network to estimate dental age. The experiments are carried out on a dataset composed of 1,636 samples. It was also experimentally confirmed that this latest feature set makes the dental age estimation more reliable for sample OPG images of ± 1-Year-old individuals' sample OPG images.

Kim et al. (56) applied a deep learning algorithm, based on CNN neural, to X-ray images of 9,435 cases (4,963 male, 4,472 female). Data in this cohort was sorted into three-age gatherings. Their study suggests that deep learning algorithms based on CNN neural networks show that the proposed approach functions evaluated based on a database of panoramic dental radiographs and worked well for accuracy.

Banar et al. (57) also implemented a fully automatic method leveraging the total capacity of the deep learning approach on their study of a dataset of 400 OPGs x-ray images with 20 OPGs per category and per gender. To overcome the

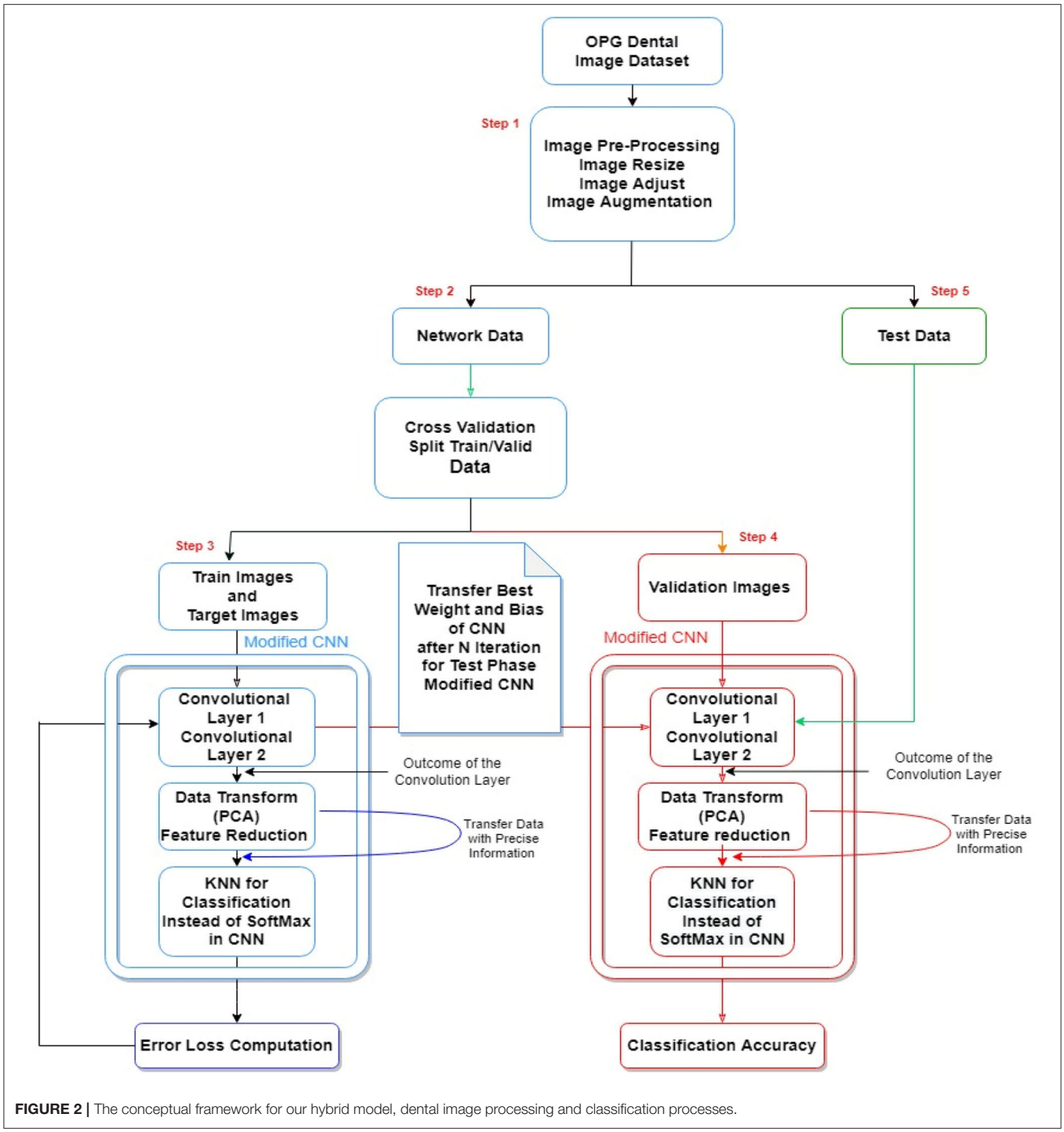


FIGURE 2 | The conceptual framework for our hybrid model, dental image processing and classification processes.

limitations of having a dataset with a limited number of embodiments, the Barnar group employed a transfer learning approach using pre-trained CNNs, and data augmentation. Their study significantly reduced the imaging assessment time over the current conventional method. For example, they have reported that the entire automatic workflow took 2.72 s to compute on average. Given the small size of the dataset, this pilot study indicated the strength of transform learning and suggested a completely automated solution capable of demonstrating outcomes not inferior to manual assessing the images.

Tuan et al. (58) introduced a semi-supervised fuzzy clustering algorithm for pre-image processing and segmentation of dental X-ray images. Their study showed that their recommended work

TABLE 1 | A display of the parameters used in the CNN network.

	Type	Filter size	Stride	# Filters	FC	Input
Layer1	Convolution	5×5	2×2	16	-	224×224×3
Layer2	Maxpooling	2×2	2×2	-	-	224×224×16
Layer3	Convolution	5×5	2×2	32	-	112×112×16
Layer4	Maxpooling	2×2	2×2	-	-	112×112×32
Layer5	Fully Connected	-	-	-	*	59×59×32

**9 for Dataset on Year, 18 for a dataset on ± 6-month, 36 for a dataset on ± 3 month and 108 for a dataset on ± 1 month.*

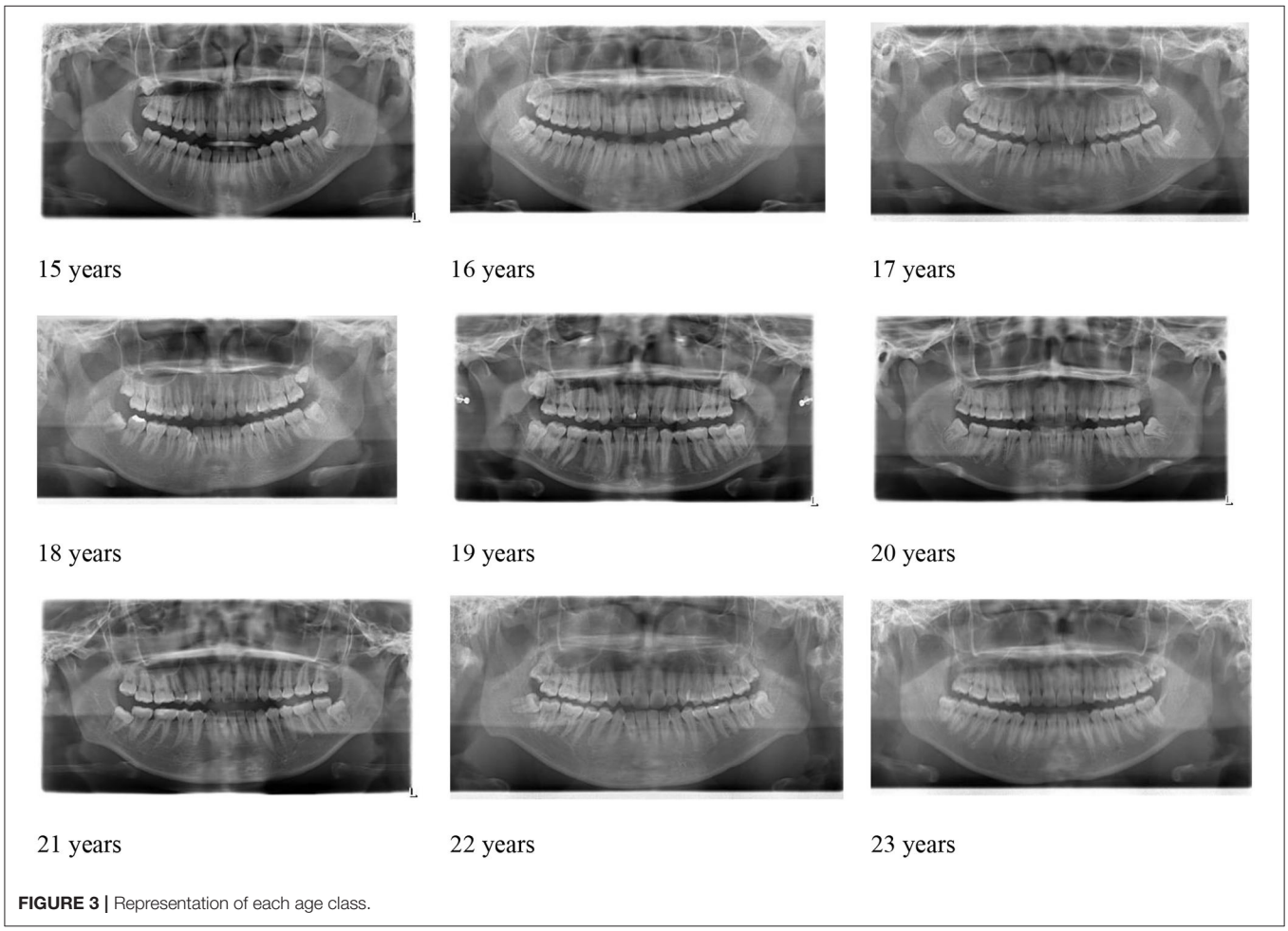


FIGURE 3 | Representation of each age class.

has superior accuracy than the initial semi-supervised fuzzy clustering and several related approaches.

In the Department of Dentistry and Study of University Sains Islam Malaysia, an age evaluation approach was tested on Malaysian adolescents between the ages of 1 to 17. Initially, the first to the third teeth were segmented, and the invariant deformation characteristics were then collected based on a deep learning method. The designed DCNN model was used to extract a broad range of features in the hierarchical layers, for example, invariance of size, rotation and deformation for sample OPG images of ± 1-Year-old individuals.

MATERIALS AND METHODS

Image Pre-processing and Data Augmentation

Prior to utilizing images in our model, we performed image processing, adjusted the contrast and highlighted the images, removed extra margins (cropping), and normalized the image pixel values at edges from zero to one. In this study, data augmentation was performed by mirroring and data duplication with rotation.

Data were classified before, Splitting for Training the model. To perform the classification, the data in the dataset were divided

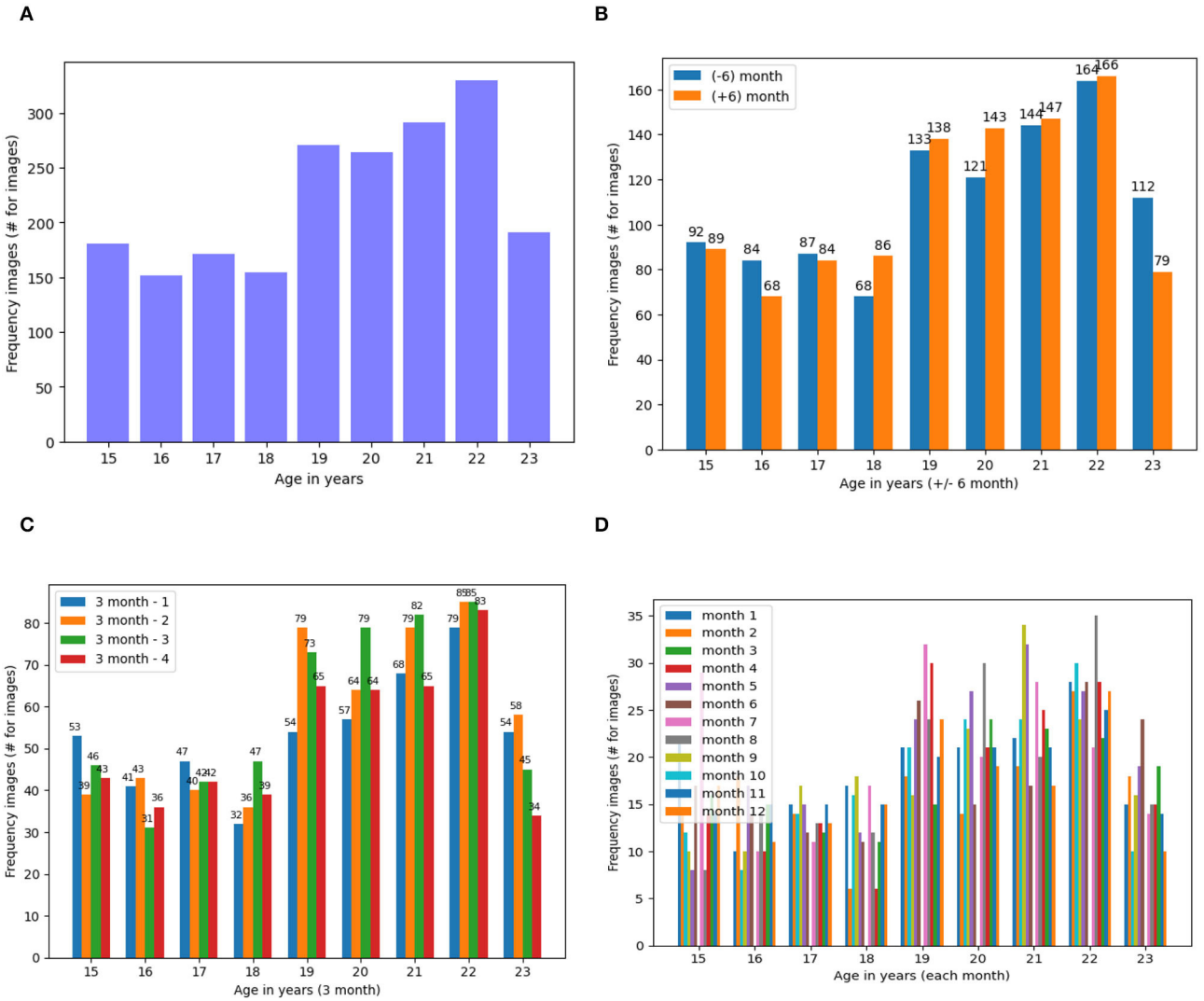


FIGURE 4 | Frequency of Dataset Images Based on the Four States of the Dataset **(A)** Dataset on Year, **(B)**. On ± 6-Month, **(C)**. On ± 3 Month and **(D)**. On ± 1 Month. **(A)** Frequency of dental images for years dataset. **(B)** Frequency of dental images for ± 6 months dataset. **(C)** Frequency of dental images for ± 3 months dataset. **(D)** Frequency of dental images for ± 1 months dataset.

TABLE 2 | Experimental result on a dental dataset based on CNN.

		Accuracy %			
Dataset	**Train/Valid**	**Dataset on year**	**Dataset on ± 6 month**	**Dataset on ± 3 month**	**Dataset on ± 1 month**
Original	Train Accuracy	77.85	76.24	72.41	70.26
	Valid Accuracy	58.91	43.72	38.75	23.29
Augmented-pre-processed	Train Accuracy	98.84	98.58	98.75	96.99
	Valid Accuracy	97.43	98.44	98.13	95.62

into two categories: train data and validation data. The cross-validation was performed with 50–50, 70–30, and 20–80 train and validation split. To take an example, 20–80 split the 20% of the data for validation and 80% for training the model. Obtaining acceptable results on various test data will indicate the correct performance of the proposed model. Thus, to further enhance

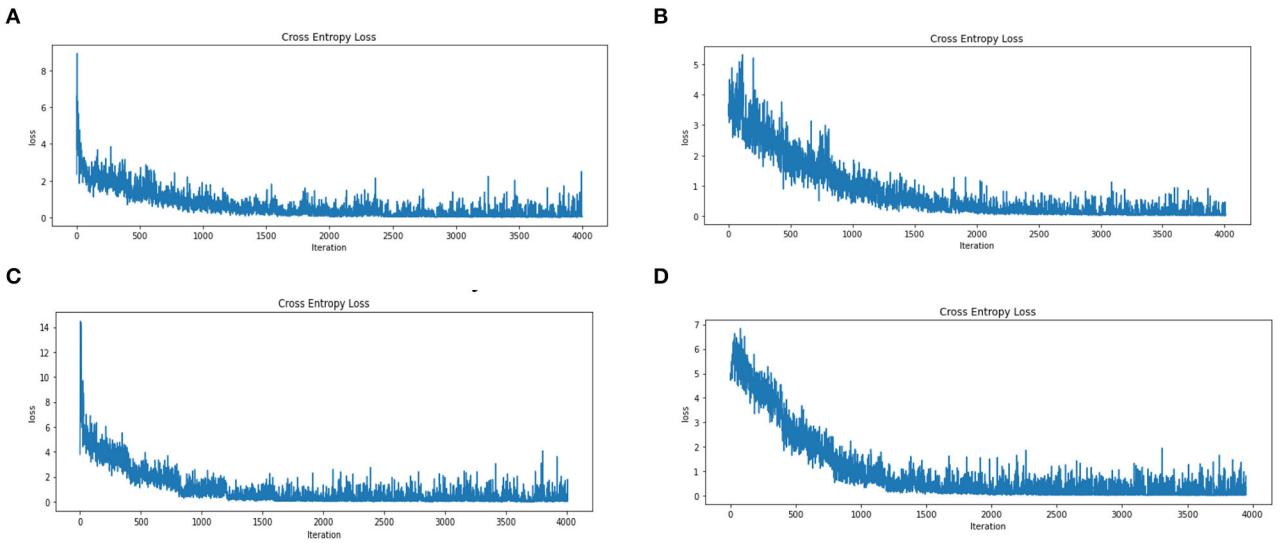


FIGURE 5 | Error Loss for Training on A Dataset Based on CNN Model on the Four States of the Augmented Dataset **(A)**. Dataset on Year **(B)**. On ± 6-Month **(C)**. On ± 3 Month and **(D)**. On ± 1 Month. **(A)** Error loss for dataset years. **(B)** Error loss for the dataset with ± 6 months. **(C)** Error loss for the dataset with ± 3 months. **(D)** Error loss for the dataset with ± 1 months.

the validation of the model, we perform additional validation of the trained model using a data set that the model was seeing for the first time. The additional validation dataset included various features such as OPGs of different races, with different genders and at different ages.

The Architecture of HCNN-KKN Proposed Model

In this study, we proposed a model based on a combination of the HCNN-KKN. Steps in our hybrid model depicted in **Figure 2** for pre-image processing and the dental classification are as follows:

The architecture of the proposed HCNN-KNN model included an inputs layer, two Convolutional layers, a fully connected layer, a PCA transform data layer, and a KNN layer for classification (**Figure 1**). Using RGB images (224 x 224 pixels) as input, each image was fed to a convolutional layer (i.e., the second layer) containing 16 convolutional filters and a kernel size of 5 and 2 paddings. Low-level features like edges, blobs, shapes were obtained. In addition to the 16 convolutional normalization filters, batch normalization (16 layers), followed by a non-linear activation function via rectified linear unit (ReLU) were performed. The convolutional layer is termed "conv1". Maxpooling layer downsampled the model with stride 2 and 2 paddings.

As a result, the Maxpooling layer reduced the dimensions of the images to 112 × 112 pixels, and the final output of this layer is equal to 112 × 112 × 16. These images were fed to the second convolutional block with a similar structure to the first. The later conventional block has 32 filters with 5 and 2 paddings kernel, resulting in the 56 × 56 × 32 images output. In the following step, all the fully connected layers were used to connect all the neurons in different layers (**Table 1**). To reduce the less important features to the next layer in the CNN model, the layers are connected with a 50% dropout between fully connected layers. We then applied a classification approach on a cohort of individuals between 15 to 23 years old and stratified all the cases with intervals of ± 6-months, ± 3-months, and ± 1-month. This resulted in 9 categories for each year (+_ Year), 18 categories for the 6 months intervals (each 6 months), 36 categories for each trimester of each year (each 3 months), 108 for every 12 months from each year (each 1 month). This resulted in the last fully connected layer to modify nine classification tasks with parameters like bias learn rate factor and weight learn rate factor. **Figure 1**: Illustration of the CNN Network elements.

In this step, a combined CNN model with KNN is used as a classifier to evaluate the class event instead of the SoftMax probability layer. Thus, before the data enters a fully connected layer of KNN layers, we proposed that in this model, the data were transferred in the space with more precise information and smaller dimensions using the principal component analysis method so-called "Fully_PCA". Therefore, fully connected layer neurons after PCA data transformation and reductions are considered overlays of the KNN classification layer. In this case, the KNN input is assumed to be 500.

The HCNN-KNN model algorithm pseudo-code proposed in this bone age study is based on dental images. In the KNN network, the distance metric used is the Euclidean Distance metric and the "Ks" are 1, 2, 3. The model is trained using an SGD momentum optimiser with a 0.0001 learning rate. The model is then trained for 10 epochs with a mini-batch size of 16 images. After learning the network and classification operations, the error loss rate is obtained to evaluate the proposed network model in subsequent epochs of network learning.

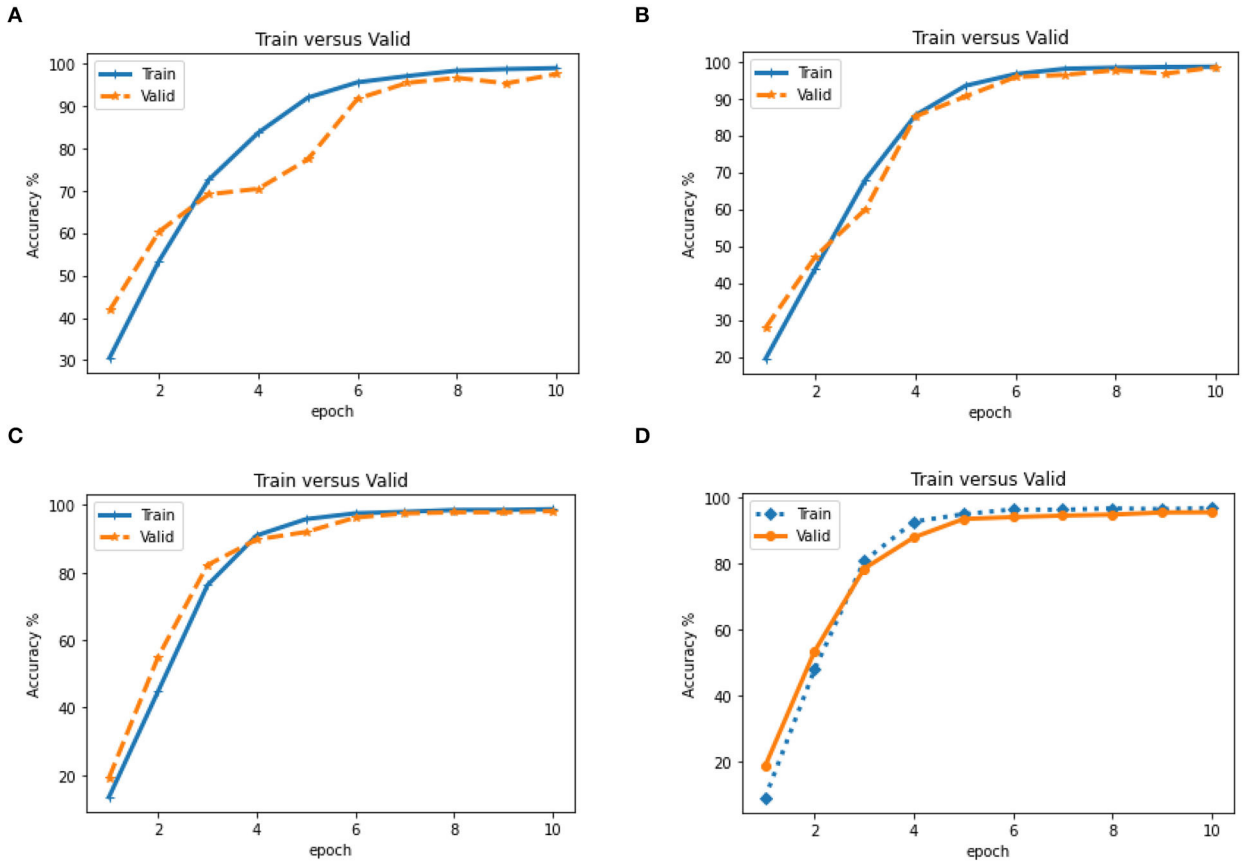


FIGURE 6 | Evaluation of accuracy (Train vs. Valid) in different epoch based on CNN model on four states of the dataset **(A)**. Dataset on Year **(B)**. on ± 6-month **(C)**. on ±3 month and **(D)**. on ± 1 month. **(A)** Train vs. Valid for dataset years. **(B)** Train vs. Valid for the dataset with ± 6 months. **(C)** Train vs. Valid for the dataset with ± 3 months. **(D)** Train vs. Valid for the dataset with ±1 months.

Evaluating the Proposed Model on Validation Data

The network neurons' optimal weight and bias were evaluated following the network training step. The validation data, segmented according to **Figure 2** (step 2), was then categorized on the CNN network with optimal parameters. The components of the classification algorithms were determined according to the training phase process and the criteria that allow the data to pass through the CNN-PCA-KNN layers. Finally, the identification rate was evaluated in which accuracy of the index was selected according to the Confusion Metrics. In brief, the sum of TP + TN (TP: the number of times the class is upbeat and correctly detected, FP: the number of times the class was positive and undiagnosed), divided with the sum of TP + FP + TN + FN. TP is the number of times the class is positive and correctly detected, FP abbreviation for the number of times the class was positive and undiagnosed. The TN, denotes the number of times the class is incorrect (negative) and correctly diagnosed. The FN represents the number of times each class is erroneous (negative) or not correctly recognized.

Validating the Proposed Model on a Test Dataset

To further validate the accuracy of the proposed model, we used test data which the model had not seen before, and was not part of the training dataset. This dataset was collected separately from the trained dataset and had different properties. This data combines X-ray OPG images of other races, with different genders and at different ages. This data enters the CNN network as a validation step and is categorized after passing through conv1, pooling1, conv2, pooling2, fully connected layer, PCA layer, and KNN layer. A vital feature of this step is the validation of the proposed model. Achieving acceptable results on various test data will indicate the satisfactory performance of the proposed model.

Computational Hardware Requirement

For this study, we used an NVIDIA GeForce GTX 1080 8GB, 8GB memory, Intel Core i7, and 3.40GHz. We also used Python libraries NumPy, Matplotlib, Sklearn, Metrics, PyTorch, Torchvision to perform the analyses.

RESULT AND DISCUSSION

We discussed the database used in this article, the tools, and the architecture introduced in the following. Finally, we evaluated our proposed model on the dataset.

Datasets Description

Dental OPG is a panoramic radiograph that scans the upper and lower jaws with a two-dimensional view that shows a semicircle from ear to ear. The used images in this study are dental OPG data, collected from dental teaching institutes and private dental clinics in Malaysia. These collected data with the different image sizes are resized before importing into the proposed model. The original image size for all data has been changed to 600 × 1,024 pixels.

Subjects from this study were randomly selected from 1,922 patients between 15 and 23 years of age, and this age range determines the age of minors which could help the model find more specific features and train the model more accurately, particularly for smaller time intervals. Moreover, to further validate the model, we used a dataset that hasn't been presented to the model during conventional training or the initial validation. Using an additional dataset including 130 random images, we showed that pre-training the model on such a prominent age recognition features (i.e. shorter time intervals, form of age range by year, age range by month, by season (3 months), and 6 months (Dataset on Year, on ± 6 months, on ± 3 months, and ± 1 month), allows us to evaluate our proposed method on a test database to enhance the performance of the proposed method (further details are provided in the experimental Result, **Table 1**, **Figure 3**).

We elucidated the frequency of the dataset generated as the result of data augmentation when augmented more than 150, 68, 32 and 7 images in 1 year, 6 months, 3 months and 1-month categories, respectively (**Figure 4**).

CNN Model Initial Accuracy on Four Different States of the Dataset

As the first experiment, we applied a CNN model on two other dataset states. In stage one, training and validation accuracy were evaluated from the original dataset without any pre-processing and data augmentation. In stage two, training and validation accuracy were assessed post-pre-processing operations on the data augmentation operations (**Table 2**).

We observed that pre-processing and data augmentation operations consistently improved the accuracy by approximately 20% and 40% for training and validation, respectively (**Table 2**).

We then evaluated the loss reduction for each step (i.e., four separate modes) of this data set (1 Year, ± 6 months, ± 3 months, and ± 1 months) with 4,000 repetitions (i.e., epoch equivalent to 10 with about 400 batch size 16). Note that the results are obtained on cross-validation with 80% training and 20% validation and segmentation on dataset images. We observe a sharp decay in loss reduction values for both the 1 year and ± 3 months dataset, while the ± 6 months and ± 1 month dataset showed a more gradual decay (**Figure 5**).

We also evaluated the accuracy of the proposed model on train and validation data in 10 epochs. We observed that the accuracy of train data is generally slightly higher than validation data (**Figure 6**). Therefore, the upward trend inaccuracy at 10 epochs indicates an increase in CNN learning.

Combining HCNN-KNN Enhanced the Accuracy of the Model

We used a combined methodology of convolutional neural networks and the nearest neighbor (classification) to further enhance the model's accuracy. The resulting hybrid model (HCNN-KNN) contained three main CNN, PCA, and KNN layers. Our findings showed that the KNN classification with 500 features selected from "Fully_PCA" (PCA data transform after fully connected in CNN) had increased the accuracy of the original data to approximately the same level of augmented data to 99.98, 99.96, 99.87, and 98.78 for 1 year, ± 6-months, ± 3-months, and on ± 1-month dataset, respectively. Intrinsically, we observed accuracy for train data to be equal to 100% (**Table 3**). This outcome indicated is fully trained using the training dataset. It is noteworthy that the results obtained for k-nearest neighbors value k = 1 in the proposed HCNN-KNN hybrid model. We have also evaluated the KNN method in the proposed HCNN-KNN hybrid model with a "k" value of 2 and 3 (**Table 4**). Model validation is essential for building a model and is susceptible to common pitfalls when splitting the training dataset (59, 60). To evaluate the possibility of overfitting pitfall resulting from splitting the training dataset, particularly for small ks (e.g., $k = 1$–3), we further validate the model using a dataset of dental images separate from the data used for initial training and validation; contained its special features. The unique features of these images are related to different races, different genders and are selected in the age range of 15 to 23. While obtaining acceptable results on these data indicates the proper performance of the KNN classifier even with k = 1.

Evaluation of Different Cross-Validation

Different cross-validation of datasets was applied to evaluate our method (e.g., 50–50, 70–30 and 20–80 train-valid cross-validation, **Table 5**).

Evaluation of Proposed Model on Second Dataset (Test Dataset)

One of this study's innovative and key features is the additional validation step in which a set of images from a diverse cohort has been used to further validate the proposed model. After the initial training and validation, the images are only available to the model. The dataset had a combination of 130 X-Ray OPG images with different races (e.g., Malay, Indian and Chinese), with distinct features, various age ranges and normal distribution of ordinary dentistry, orthodontic and malignant images (**Figure 7**).

As **Table 6** shows, the proposed HCNN-KNN model has obtained better results on the test data. Getting satisfactory results on a new dataset outside the dataset images indicates the performance of the proposed model. Thus, it seems that the proposed HCNN-KNN model has obtained superior results on the tested dataset. Obtaining satisfactory results on a new

TABLE 3 | Experimental result on a dental dataset based on HCNN-KNN.

		Accuracy %			
Dataset	**Train/Valid**	**Dataset on Year**	**Dataset on ± 6 month**	**Dataset on ± 3 month**	**Dataset on ± 1 month**
Original	Train Accuracy	100	100	100	100
	Valid Accuracy	99.93	99.84	99.74	98.15
Augmented	Train Accuracy	100	100	100	100
	Valid Accuracy	99.98	99.96	99.87	98.78

TABLE 4 | Accuracy results of K = 1, 2, 3 on the validation data.

	Accuracy %			
K	**Dataset on Year**	**Dataset on ± 6 month**	**Dataset on ± 3 month**	**Dataset on ± 1 month**
K = 1	99.98	99.96	99.87	98.78
K = 2	59.26	55.61	52.64	51.20
K = 3	43.89	41.64	38.02	33.77

TABLE 5 | Evaluation of accuracy on different cross-validation on a dataset.

			Accuracy on different Cross-Validation (valid-Train)		
Dataset Type	**Methods**	**Train/Valid**	**20-80**	**30-70**	**50-50**
Dataset on Year	CNN	**Train**	98.84	98.16	97.20
		Valid	97.43	94.08	85.58
	HCNN-KNN	**Train**	**100.00**	**100.00**	**100**
		Valid	**99.98**	**99.97**	**99.95**
	Resnet (Deep model)	**Train**	99.04	98.83	98.80
		Valid	98.25	95.75	86.03
Dataset on ± 6 month	CNN	**Train**	98.58	98.50	96.75
		Valid	98.44	94.72	86.66
	HCNN-KNN	**Train**	**100.00**	**100.00**	**100**
		Valid	**99.96**	**99.78**	**99.78**
	Resnet (Deep model)	**Train**	98.66	98.53	98.42
		Valid	98.69	95.47	87.43
Dataset on ± 3 month	CNN	**Train**	98.75	98.48	97.77
		Valid	98.13	96.25	84.61
	HCNN-KNN	**Train**	**100.00**	**100.00**	**100.00**
		Valid	**99.87**	**99.82**	**99.68**
	Resnet (Deep model)	**TrainValid**	98.69 98.44	98.7496.42	98.70 86.23
Dataset on ± 1 month	CNN	**Train**	96.99	97.12	94.46
		Valid	95.62	92.40	80.32
	HCNN-KNN	**Train**	**100.00**	**100.00**	**100.00**
		Valid	**98.78**	**98.52**	**98.29**
	Resnet (Deep model)	Train	97.25	97.24	97.15
		Valid	95.63	93.16	90.27

Bold value is the maximum accuracy compared to the others results.

dataset outside the dataset images indicated the outstanding performance of the proposed model (**Table 6**). The HCNN-KNN model effects have also been tested for k 1, 2, and 3 (**Table 6**). The outcome shows the proper performance of the proposed model for k = 1. In KNN, where k is larger, the distance becomes more critical, overcoming the KNN principle that closer

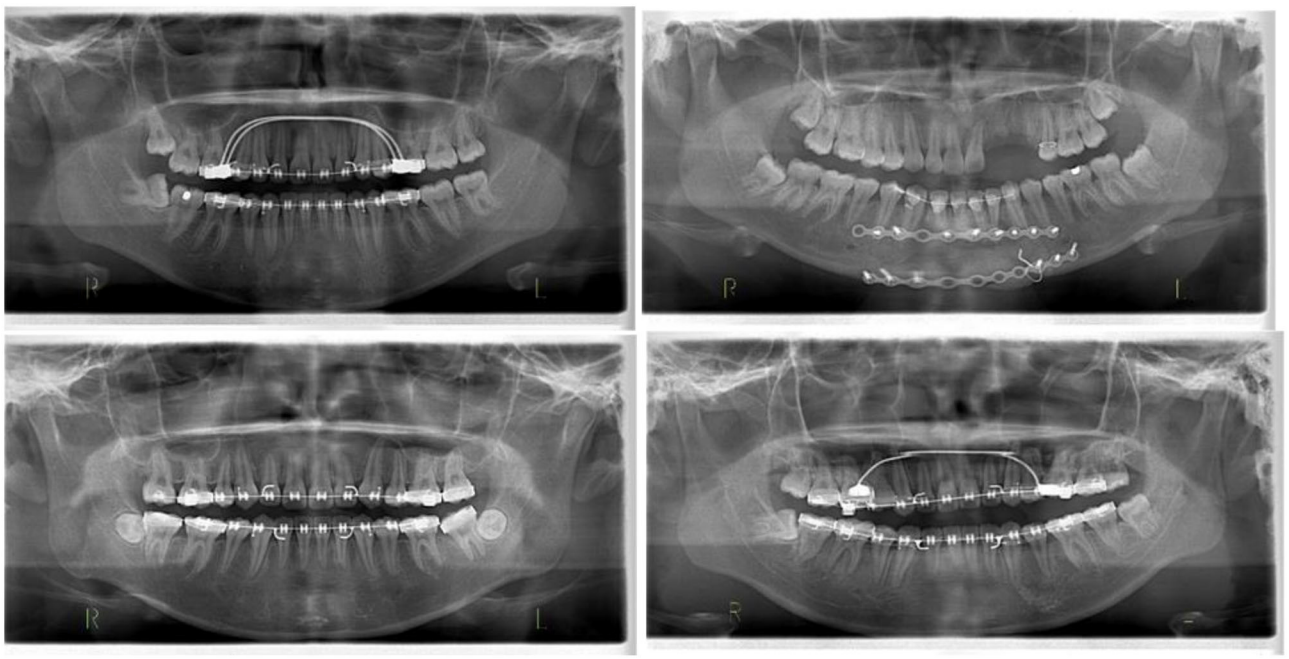


FIGURE 7 | Different images of test data to evaluate the orthodontics (malignant) of the proposed model.

TABLE 6 | Accuracy Result of the Proposed HCNN-KNN Model on Test New Dataset.

			CNN	HCNN-KNN
All test Data	**Test data (130 dental images)**	22.93	**K = 1 99.48**	
			K = 2 74.61	
			K = 3 70.22	
Based on Different Races	**Malay (71 Images)**	25.23	**K = 1 100**	
			K = 2 76.05	
			K = 3 73.23	
	Indian (16 Images)	20.61	**K = 1 93.75**	
			K = 2 75.00	
			K = 3 68.75	
	Chinese (43 Images)	21.37	**K = 1 97.67**	
			K = 2 72.09	
			K = 3 69.76	

Bold value is the maximum accuracy compared to the others results.

TABLE 7 | Comparison between the proposed HCNN-KNN model and other studies in bam based on dental images.

References	Accuracy	Models
CNN	**97.43**	**Other model test result**
HCNN-KNN	**99.98**	**The proposed hybrid model result**
ResNet	**98.25**	**Other model test result**
GoogLeNet Inception	**55.69**	**Other model test result**
SVM	92.1	(11)
FNN-TLBO	89.00	(61)
Multilayer Perceptron	90.00	(54)

Bold value is the maximum accuracy compared to the others results.

neighbors have the same density or classes. Since the classes in our dataset are very near to each other and are pre-processed with noise reduction, the result shows acceptable accuracy at K = 1.

Figure 8 shows the confusion matrix of 97.95 % for the 18-year-old age range which was correctly estimated, and 2.04 % was incorrectly identified for the 19-year-old.

Models Accuracy

The collected data and the proposed model have been tested with other models such as ResNet, CNN, GoogLeNet Inception (**Table 7**). Our finding shows that the hybrid HCNN-KNN model has obtained much higher accuracy than different conventional classification algorithms tested.

The novelty of our study is that to the best of our knowledge, no other study reported measurements of ± 6 months, ± 3 months, and ± 1 month so far, but only compared by the year. However, the results of this proposed model are compared with previous studies by years of accuracy. For example, Avuçlu used the Multilayer Perceptron model with 1,315 dental OPG images, and the highest accuracy was 90% just for accuracy of the years. Also, in 2020, Hussein used the SVM algorithm for a population of 1,429 dental radiographs, and they only achieved 92% accuracy for years. Therefore, based on **Table 7**, the results

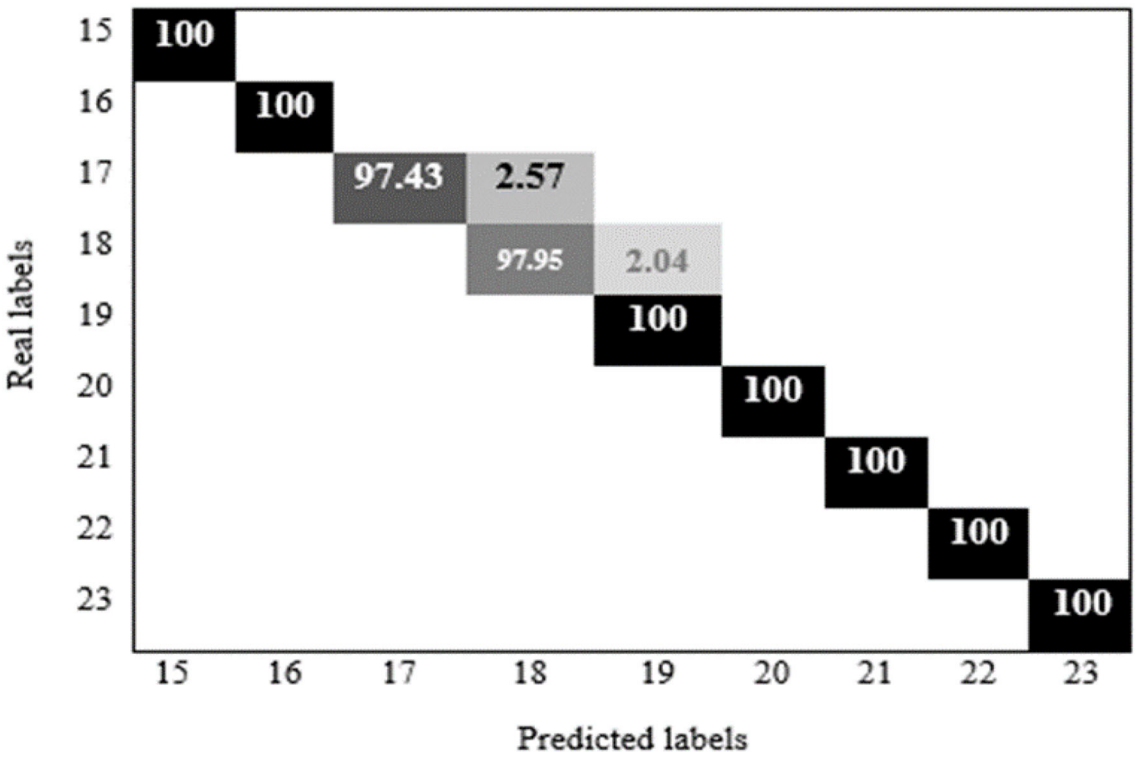


FIGURE 8 | Confusion matrix of hybrid HCNN-KNN proposed model application on enhancement dataset.

of the comparative studies for accuracy by years are still below the accuracy achieved in this study which is 99.98.

CONCLUSIONS

For the first time, with precision to the range of +/- 6 months, this novel implementation developed an HCNN-KNN model, used for BAM to increase the existing model's accuracy and prevent CNN overload situations. We considered more specific cases of bone age measurement with the help of Dental X-Ray OPG images to solve the problem of bone age measurement in determining the age range of 15 to 23 years based on the Year, ± 6 months, ± 3 months, and ± 1 month. We integrated the methodology of convolutional neural networks for extraction and analysis of information and features in dental images) and the nearest neighbor (classification method). The primary purpose of this proposed model is to use KNN instead of SoftMax in the fully connected layer to increase the performance of convolutional networks in the classification phase. Using principal component analysis as data transform and feature dimensionality reduction in a fully connected layer before classifying KNN as Fully_PCA. The primary purpose of this method is to transfer data to space with more specific information to increase the performance of the classification phase.

Our proposed model achieved the accuracy of 99.98, 99.96, 99.87, and 98.78 in 1 year, ± 6-month, ± 3-month, and ± 1-month range, respectively. Our proposed method evaluated different cross-validation of the dataset. 50–50 train-valid, 70–30 train-valid and 20–80 train-valid cross-validation on the dataset. Evaluating the proposed model on a new dataset with different races also proved the superior performance of the model. The benchmarking with current existing models also showed that the HCNN-KNN model is the best model for bone age measurement.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

FS and SB designed the study. FS wrote the manuscript, collected the data, carried out the analyses, including statistical analyses and the implementation of machine learning methods and generated all figures and tables. FS, SB, PO, and HA-R edited the manuscript. HA-R and PO were not involved in any analysis. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank for the support of Taif University Researchers Supporting Project number (TURSP-2020/239), Taif University, Taif, Saudi Arabia. We are thankful to the computer science and engineering school, SCE Taylor's University, for providing the scholarship to complete the studies. HA-R has been supported by the UNSW Scientia Program Fellowship.

REFERENCES

1. Mohammad N, Yusof MYPM, Ahmad R, Muad AM. Region-based segmentation and classification of mandibular first molar tooth based on Demirjian's method. *J Phys.* (2020) 1502:012046. doi: 10.1088/1742-6596/1502/1/012046
2. Kim S, Lee YH, Noh YK, Park FC, Auh Q. Age-group determination of living individuals using first molar images based on artificial intelligence. *Sci Rep.* 11:1–11. (2021) doi: 10.1038/s41598-020-80182-8
3. Wik EH, Martínez-Silván D, Farooq A, Cardinale M, Johnson A, Bahr R. Skeletal maturation and growth rates are related to bone and growth plate injuries in adolescent athletics scand. *J Med Sci Sports.* (2020) 30:894–903. doi: 10.1111/sms.13635
4. McGill D. Bone Age. *Endocrine Conditions in Pediatrics.* Cham: Springer. (2021).
5. Dahlberg PS, Mosdøl A, Ding Y, Bleka Ø, Rolseth V, Straumann GH, et al. systematic review of the agreement between chronological age and skeletal age based on the greulich and pyle atlas. *Eur Radiol.* (2019) 29:2936–48. doi: 10.1007/s00330-018-5718-2
6. Gurpiner F, Kaya H, Dibeklioglu H, Salah AA. Kernel ELM and CNN based facial age estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.* Las Vegas, NV: IEEE (2016). p. 80–6.
7. Mohammad N, Muad AM, Ahmad R, Yusof MYPM. Reclassification of Demirjian's mandibular premolars staging for age estimation based on semi-automated segmentation of deep convolutional neural network. *Forensic Imag.* (2021) 24:200440. doi: 10.1016/j.fri.2021.200440
8. Alkaabi S, Yussof S, Al-Mulla S. Evaluation of convolutional neural network based on dental images for age estimation. *ICECTA.* (2019) 2019:1–5. doi: 10.1109/ICECTA48151.2019.8959665
9. Ezhil I, Jagannathan N, Kumar MP. Estimation of age from physiological changes of teeth. *Drug Invent Today.* (2018).
10. Cular L, Tomaić M, Subašić M, Šarić T, Sajković V, Vodanović M. Dental age estimation from panoramic X-ray images using statistical models. In: *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis.* Ljubljana: IEEE (2017). p. 25–30. doi: 10.1109/ISPA.2017.8073563
11. Houssein EH, Mualla N, Hassan M. Dental age estimation based on X-ray images. *Computers Materials Continua.* (2020) 2:591–605. doi: 10.32604/cmc.2020.08580
12. Shi W, Yan G, Li Y, Li H, Liu T, Sun C, et al. Fetal brain age estimation and anomaly detection using attention-based deep ensembles with uncertainty. *Neuroimage.* (2020) 223:117316. doi: 10.1016/j.neuroimage.2020.117316
13. Alinejad-Rokny H. Proposing on Optimized Homolographic Motif Mining Strategy Based on Parallel Computing for Complex Biological Networks. *J Med Imaging & Health Infor.* (2016) 6:416–24. doi: 10.1166/jmihi.2016.1707
14. Javanmard A, Montanari A. IMA. State evolution for general approximate message passing algorithms, with applications to spatial coupling Information and Inference: A. *J IMA.* (2013) 2:115–44. doi: 10.1093/imaiai/iat004
15. Ahmadinia M, Meybodi MR, Esnaashari M. Energy-efficient and multi-stage clustering algorithm in wireless sensor networks using cellular learning automata. *IETE J Res.* (2013) 59:774–82. doi: 10.4103/0377-2063.126958
16. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi AJ SS. Band AEJ Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries alex. *Eng J.* (2021) 60:457–64. doi: 10.1016/j.aej.2020.09.013
17. Hosseinpoor M. B-Koura IO, Yahia AJC, Composites C. New methodology to evaluate the reynolds dilatancy of self-consolidating concrete using 3D image analysis-coupled effect of characteristics of fine mortar and granular skeleton. *Cem Concr Compos.* (2020) 108:103547. doi: 10.1016/j.cemconcomp.2020.103547
18. Esmaeili L, Minaei-Bidgoli B, Alinejad-Rokny H, Nasiri M. Hybrid recommender system for joining virtual communities. *Res J Appl Sci Eng Technol.* (2012) 4:500–509.
19. Hussein IJ, Burhanuddin MA, Mohammed MA, Benameur N, Maashi MS. Maashi MS. Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (HOG). *Expert Systems.* (2021) 1:e12789. doi: 10.1111/exsy.12789
20. Khan Ah Iskandar DA, Al-Asad JF, El-Nakla S, Alhuwaidi SA. Statistical feature learning through enhanced delaunay clustering and ensemble classifiers for skin lesion segmentation and classification. *J Theor Appl Inf Technol.* (2021) 99:5.
21. Mahmoudi MR, Akbarzadeh H, Parvin H, Nejatian S, Rezaie V. Consensus function based on cluster-wise two level clustering. *Artificial Intelligence Review.* (2021) 54:639–65. doi: 10.1007/s10462-020-09862-1
22. Niu H, Khozouie N, Parvin H, Beheshti A, Mahmoudi MR. An ensemble of locally reliable cluster solutions. *Appl Sci.* (2020) 10:1891. doi: 10.3390/app10051891
23. Parvin H, Minaei B, Alizadeh H, Beigi A. A novel classifier ensemble method based on class weightening in huge dataset. *International Symposium on Neural Networks.* (2011) 2011:144–50. doi: 10.1007/978-3-642-21090-7_17
24. Parvin H, Parvin S. A classifier ensemble of binary classifier ensembles. *IJLMS.* (2013) 1:37–47. doi: 10.12785/ijlms/010204
25. Parvin H, MirnabiBaboli M, Alinejad-Rokny H. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Eng Appl Artif Intell.* (2015) 1:34–42. doi: 10.1016/j.engappai.2014.08.005
26. Alinejad-Rokny H, Pourshaban H, Orimi AG. Baboli Network motifs detection strategies and using for bioinformatic networks. *J Bionanoscience.* (2014) 8:353–9. doi: 10.1166/jbns.2014.1245
27. Bahrani P, Minaei-Bidgoli B, Parvin H, Mirzarezaee M, Keshavarz A. User and item profile expansion for dealing with cold start problem. *J Intell Fuzzy Syst.* (2020) 38:4471–83. doi: 10.3233/JIFS-191225
28. Sharifrazi D, Alizadehsani R, Hassannataj Joloudari J, Shamshirband S, Hussain S, Alizadeh Sani Z, et al. CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering. *Math Biosci Eng.* (2021) 3:2381–402. doi: 10.20944/preprints202007.0650.v1
29. Dental age estimation in East Asian population with least squares regression. In: *International Conference on Advanced Machine Learning Technologies and Applications.* Cham: Springer (2018). p. 653–60.
30. Farhadian M, Salemi F, Saati S, Nafisi N. Dental age estimation using the pulp-to-tooth ratio in canines by neural networks. *Imaging Sci Dent.* (2019) 49:19–26. doi: 10.5624/isd.2019.49.1.19
31. Alinejad-Rokny H, Anwar F, Waters S, Davenport M, Ebrahimi D. Source of CpG depletion in the HIV-1 genome. *Mol Biol Evol.* (2016) 33:3205–12. doi: 10.1093/molbev/msw205
32. Alinejad-Rokny H, Sadroddiny E, Scaria V. Machine learning and data mining techniques for medical complex data analysis. *Neurocomputing.* (2018) 276:1. doi: 10.1016/j.neucom.2017.09.027
33. Bayati M, Rabiee HR, Mehrbod M, Vafaee F, Ebrahimi D, Forrest AR, et al. a user-friendly and robust tool for identification and classification of mutational signatures and patterns in cancer genomes. *Sci Rep.* (2020) 10:1–11. doi: 10.1038/s41598-020-58107-2
34. Dashti H, Dehzangi A, Bayati M, Breen J, Lovell N, Ebrahimi D. Integrative analysis of mutated genes and mutational processes reveals seven colorectal cancer subtypes. *BMC Bioinformatics.* (2020) 23:1–24. doi: 10.1186/s12859-022-04652-8
35. Ghareyazi A, Mohseni A, Dashti H, Beheshti A, Dehzangi A, Rabiee HR. Whole-genome analysis of *de novo* somatic point mutations reveals novel mutational biomarkers in pancreatic cancer. *Cancers.* (2021) 13:4376. doi: 10.3390/cancers13174376
36. Hosseinpoor M, Parvin H, Nejatian S, Rezaie V, Bagherifard K, Dehzangi A. Proposing a novel community detection approach to identify cointeracting genomic regions. *Math Biosci Eng.* (2020) 3:2193–217. doi: 10.3934/mbe.2020117
37. Javanmard R, JeddiSaravi K. Proposed a new method for rules extraction using artificial neural network and artificial immune system in cancer diagnosis. *J Bionanoscience.* (2013) 7:665–72. doi: 10.1166/jbns.2013.1160
38. Rajaei P, Jahanian KH, Beheshti A, Band SS, Dehzangi A, VIRMOTIF. A user-friendly tool for viral sequence analysis. *Genes.* (2021) 12:186. doi: 10.3390/genes12020186
39. Alinejad-Rokny H, Pedram MM, Shirgahi H. Discovered motifs with using parallel Mprefixspan method. *Sci Res Essays.* (2011) 20:4220–6. doi: 10.5897/SRE11.212
40. Ahmadinia M, Alinejad-Rokny H, Ahangarikiasari H. Data aggregation in wireless sensor networks based on environmental

similarity: A learning automata approach. *J Networks.* (2014) 9:2567. doi: 10.4304/jnw.9.10.2567-2573
41. Parvin H, Asadi M. An ensemble based approach for feature selection. *Res J Appl Sci.* (2011) 9:33–43.
42. Niu H, Xu W, Akbarzadeh H, Parvin H, Beheshti A. Deep feature learnt by conventional deep neural network. *Comput Electr Eng.* (2020) 84:106656. doi: 10.1016/j.compeleceng.2020.106656
43. Hemalatha B, Rajkumar N, A. modified machine learning classification for dental age assessment with effectual ACM-JO based segmentation. *Int J Bio-Inspir Com.* (2021) 2:95–104. doi: 10.1504/IJBIC.2021.114089
44. Ouali Y, Hudelot C, Tami M. An overview of deep semi-supervised learning. *arXiv[Preprint].* (2020). arXiv: 2006.05278. Available online at: https://arxiv.org/pdf/2006.05278.pdf
45. Goldberg AB. *New Directions in Semi-Supervised Learning.* Doctoral dissertation. Madison : University of Wisconsin (2010).
46. Spampinato C, Palazzo S, Giordano D M. Aldinuccii RJM. A Leonardi Deep learning for automated skeletal bone age assessment in X-ray images. *Medical Image Analysis.* (2017) 36:41–51. doi: 10.1016/j.media.2016.10.010
47. Kalantari A, Kamsin A, Shamshirband S, Gani A H. Alinejad-Rokny A. Chronopoulos TJN computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions. *Neurocomputing.* (2018) 276:2–22. doi: 10.1016/j.neucom.2017.01.126
48. Shamshirband S T. Rabczuk, and K-W, Chau JIA. A survey of deep learning techniques: application in wind and solar energy resources. *IEEE Access vol.* (2019) 7:164650–66. doi: 10.1109/ACCESS.2019.2951750
49. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT. A review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues. *J Biomed Inform.* (2021) 113:103627. doi: 10.1016/j.jbi.2020.103627
50. Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, et al. Jadoon JIA. A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access.* (2019) 7:150530–9. doi: 10.1109/ACCESS.2019.2947484
51. Li J, Koyamada S, Ye Q, Liu G, Wang C, Yang R, et al. Suphx: Mastering mahjong with deep reinforcement learning. *arXiv[Preprint].* (2020). arXiv: 2003.13590. Available online at: https://arxiv.org/pdf/2003.13590.pdf
52. Sharifonnasabi F, Jhanjhi N, John J, Alaboudi A, Nambiar P. review on automated bone age measurement based on dental OPG images. *Int J Eng Res.* (2020) 132:5408–22.
53. Blatt SH, Shields JR, Michael AR. Dental calculus reveals life history of decedents in forensic cases: An anthropological perspective on human identification. *Forensic Genom.* (2022) 2:5–16. doi: 10.1089/forensic.2022.0003
54. Avuçlu E, Başçiftçi F. The determination of age and gender by implementing new image processing methods and measurements to dental X-ray images. *Measurement.* (2020) 149:106985. doi: 10.1016/j.measurement.2019.106985
55. Fatih H, Catalin G. Recreation and physical activity of young girls with intellectual disability. *Sci Mov Health.* (2020) 20:237–41.
56. Kim JH, Kim BG, Roy PP, Jeong DM. Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. *IEEE Access.* (2019) 7:41273–85.
57. Banar BB. *Pengembangan media pembelajaran mobile learning dengan aplikasi berbasis android untuk meningkatkan hasil belajar siswa (studi pada peserta didik kelas XI Program Keahlian Otomatisasi Tata Kelola Perkantoran SMK Negeri 1 Pogalan* (Doctoral dissertation, Universitas Negeri Malang).
58. Tuan TM. Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints. *Eng Appl Artif Intell.* (2017) 59:186–95. doi: 10.1016/j.engappai.2017.01.003
59. Wasule V, Sonar P. Classification of brain MRI using SVM and KNN classifier. pp. 218-223.
60. Greener JG, Kandathil SM, Moffat L. Jones DT. A guide to machine learning for biologists. *Nat Rev Mol.* (2022) 1:40–55. doi: 10.1038/s41580-021-00407-0
61. Hemalatha. A versatile approach for dental age estimation using fuzzy neural network with teaching learning-based optimization classification. *Multimed Tools Appl.* (2020) 79:3645–65. doi: 10.1007/s11042-018-6434-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sharifonnasabi, Jhanjhi, John, Obeidy, Band, Alinejad-Rokny and Baz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fpubh.2022.902648

Analysis of the Impact of Supply Chain Relationship Strategy on Corporate Cash Dividend Policy Based on Chinese Data

*Xingquan Yang and Lingyan Sun**

School of Economics and Management, Shihezi University, Shihezi, China

With the development and application of e-commerce in the process of supply chain integration, the choice of supplier centralized strategy or decentralized mode and how to use supplier financing have become significant contents of supply chain management. This study investigates the effect of competition and cooperation on the corporate cash dividend policy under the influence of the supplier relationship strategy and its mediating mechanism based on Chinese data. The motivation of this study is to provide a basis for enterprises to grasp the dynamic evolution process of the economic consequences of supply chain relationships based on big data and adjust the relationship strategy in time to maximize the positive effects of supplier relationships. This study considers supplier concentration and supplier financing as two dimensions to measure the supplier relationship strategy and selects the balanced panel data of Chinese A-share listed companies from 2007 to 2020 as samples by applying the Logit and Tobit model. The results demonstrate that the supplier relationship is negatively correlated with the cash dividends. The intermediary effect found that the competition effect of the supplier relationship aggravates the agency conflict of enterprises and intensifies the degree of financing constraints, and thus acts on the cash dividends of enterprises. This study expands the economic consequences of relational transactions and provides an explanation of dividend policies from the perspective of a supply chain.

Keywords: supplier concentration, supplier financing, cash dividends, agency cost, financing constraints

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Rmahammad Shafi,
Mizan Tepi University, Ethiopia
Srinivasa Rao Namburi,
Bapatla Engineering College, India

****Correspondence:***
Lingyan Sun
linda.1sunny@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *23 March 2022*
Accepted: *19 April 2022*
Published: *30 May 2022*

Citation:
Yang X and Sun L (2022) Analysis of the Impact of Supply Chain Relationship Strategy on Corporate Cash Dividend Policy Based on Chinese Data. Front. Public Health 10:902648. doi: 10.3389/fpubh.2022.902648

INTRODUCTION

With the development of the supply chain, the influence of supplier relationship strategy formed by enterprise relationships on enterprise financial decisions has gradually attracted the attention of academic circles. There are numerous and extensive nonfinancial studies on the economic consequences of supplier relationship strategy, and the primary research contents include the following components: influence factors of the supplier relationship strategy (1); supplier relationship dimension (2); the influence of the supplier relationship on the enterprise performance (3, 4), supplier relationship integration (5, 6) supplier collaboration (7–9), information sharing (10), and the related incentive and constraint mechanisms (11–13). In recent years, the research with regard to the influence of the supplier relationship on enterprise finance aspects has also attracted the attention of scholars, which primarily involves the influence of the supplier relationship on the

enterprise business credit (14–16), cash holdings (17), capital structure (18–20), transaction costs (21, 22), the choice of corporate governance (23–25), bargaining power (26), earnings management (27), financial distress, and the bankruptcy risk of enterprises (28). However, the influence of the supplier relationship strategy on cash dividend policy is rarely investigated in the extant literature. Existing studies on the impact of supply chain on shareholder wealth only discuss the perspective of customers, ignoring the impact of supplier relationships on dividends (29, 30). Actually, with the development of e-commerce, the business-to-customer model saves the cost of looking for customers and weakens the advantage of buyer's market; In contrast, there is a direct market transaction relationship between suppliers and companies, so suppliers have a greater impact on the transaction costs, investment efficiency and business decisions of enterprises than customers. Therefore, the study on how supplier relationship affects corporate financial decisions not only expands the research on the stakeholder factors of dividend policy but also provides a basis for enterprises to timely judge and adjust relevant business strategies and financial decisions based on the dynamic evolution of supplier relationship observed by big data analysis.

By summarizing the existing empirical research on cash dividend policy, scholars mainly discuss the factors affecting dividend payments from the macro and micro levels: At the micro-level, it mainly includes interest correlation (31), capital structure (32), corporate governance (33, 34), risk (35), ownership structure (36), etc. At the macro level, it mainly involves the legal environment (37), market environment (38), cultural and political factors (37), and external macro-environment (39). These factors are based on the theoretical basis of information asymmetry (40) or agency cost aspect (31, 41, 42) and the financing constraint hypothesis (43, 44). By summarizing empirical research and relevant theoretical findings, although the academic research with regard to the cash dividends has been expanding the explanation of its influencing factors from the theoretical and empirical perspectives, its underlying source can be summarized as the agency theory of enterprises and the perspective of information asymmetry. The early literature on dividend influencing factors only discussed a single aspect, until the dividend tradeoff model proved by Rozeff showed that dividend policy was the result of the joint influence of agency cost and financing constraints under information asymmetry. On the one hand, the information asymmetry between managers and shareholders causes managers to control the cash flow by reducing cash dividends at the expense of shareholders' interests, maximizing their own interests (41). Therefore, increasing cash dividends can restrain agency costs. However, the increase in the cash dividends gives rise to the higher external financing cost while restraining the agency problem, and consequently, the determination of the company's optimal dividend policy requires considering both effects. In addition, in actual corporate governance, enterprises with severe agency costs simultaneously burden financing constraints, and thus, the formulation of optimal financial decisions should be according to both agency costs and financing constraints (43). Therefore, by further expanding the theory of Rozeff (45), Chae et al. (33) conducted an empirical test by using the data of the US-listed companies and proves that the relationship between the dividend payments level and the corporate governance level depends on the relative size of agency costs and external financing costs. Based on the status quo of weak corporate governance and weak investor legal protection in China's listed companies, the problem of financing constraints and agency costs caused by the information asymmetry is serious and longstanding, which curbs corporate governance as well as financial decision making. The scholars gradually began to explore the influence of agency costs and financing constraints on the corporate cash dividend policy from the dual perspectives. Lu and Wang (34) confirmed that when companies faced agency costs and financing constraints simultaneously, companies with good corporate governance would increase the payment of the cash dividend. Subsequently, from the perspective of the cash dividend policy, Zhong and Lu (46) proved that the signal transmission of the cash dividend policy should take into account the dual restriction of financing constraints and agency costs. Yu and Liang (44) proved that China's low dividend is the result of the dual effects of agency cost and financing constraints according to the deviation degree of the actual dividend level.

The relationship between enterprises and suppliers is a dynamic process in which cooperation and competition coexist, and it is internalized into the competition and cooperation relationship. The cooperative effect of suppliers may promote information sharing through relational transactions and constitute an external supervision mechanism to increase enterprises' profitability, thus alleviating the agency cost of enterprises, and consequently, affecting cash dividends. It is also possible to increase cash dividends by easing the financing constraints of enterprises by supplying the commercial credit provided by suppliers. The competitive effect of suppliers on the cash dividend policy of enterprises is primarily manifested in the predatory practices induced by both sides of the transaction to maximize their own interests. On the one hand, the predatory practices of enterprises to suppliers are reflected by opportunism and moral hazard. Thus, the management tends to take advantage of the buyer by inducing suppliers to provide them with on-the-job consumption and personal opportunities or by facilitating participation in the board of directors, and by other covert behaviors, such as transferring corporate wealth, increasing agency costs, and reducing cash dividends. On the other hand, to maximize their own interests, suppliers with bargaining advantages rip off downstream enterprises, reduce the supply of commercial credit, increase the financing constraints of enterprises, and thus reduce cash dividends. In this scenario, this study investigates whether the supplier relationship strategy induces a significant impact on the cash dividend policy of enterprises. If the relationship exists, the central mechanism of action is attributed to the cooperative effect or competition effect.

In this study, the annual reports of companies listed from 2007 to 2020 are used to investigate the two dimensions, namely, supplier concentration and supplier commercial credit financing, and judge their impact on the corporate cash dividend policy. Subsequently, these dimensions are used to explore the mediating mechanism of supplier relationship strategy

affecting cash dividends. The test results show that the supplier concentration is negatively correlated with the cash dividend payments tendency as well as the cash dividend payments level. Supplier commercial credit financing is negatively correlated with the cash dividend payments level and the cash dividend payments intention. The mediation mechanism test supports the competitive effect of the supplier relationship strategy. The supplier relationship strategy increases the agency cost of enterprises and the financing constraint of enterprises, which reduces the dividend payments tendency and the dividend payments level. It provides an explanation of the negative dividend policy based on the perspective of the supply chain.

As the rare empirical study on the impact of supplier relationship strategy on corporate cash dividends, this study explores the cash dividend policy of enterprises from the perspective of the supplier relationship strategy, which not only enriches the relevant literature concerning the influence of the supplier relationship strategy on enterprise behavior but also provides certain enlightenment for enterprises to improve the supplier relationship strategy and optimize the cash dividend policy. This study differs from Wang's (29) study that only discusses the impact of a single level of customer relationship on cash dividends. Wang (29) believes that the financial distress hypothesis and certification hypothesis are the theoretical basis for the effect of economic consequences of customer relationships on corporate cash dividends, and finally proves that the financial distress hypothesis is valid. Different from the research of Wang (29), this article from the perspective of the supplier relationship, summarizes the root causes of its economic consequences as the cooperation effect and competition effect and proves that the supplier relationship affects the agency cost and financing constraints of enterprises through competition effect, thus acting on cash dividends, which is different from the economic consequences of financial distress in customer relationships studied by Wang (29).

The contribution and research motivation of this study can be summarized as follows: First, it can provide a basis for enterprises to grasp the status of cooperation effect and competition effect of supplier relationship in time by using big data analysis, adjust financial strategy in time to improve corporate governance level and anti-risk ability, and maximize the positive effect of supplier relationship. Second, the research conclusion proves the significance of the information effect on supplier relationships, which will promote the comprehensiveness and timeliness of big data construction and provide a reference for the decision-making of regulators and investors at the supply chain level.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

The supplier relationship primarily impacts the cash dividend policy of enterprises by using the "cooperation effect" and the "competition effect," thus generating "positive" and "negative" effects, respectively.

Supplier Relations and the Corporate Cash Dividend Policy: Cooperative Effect

The cooperative relationship is an enterprise relationship based on a long-term written contract, which can play a greater role as compared to the contract and enable both parties to share information, risks, and benefits over a long period of time (47). On the one hand, the cooperation effect of suppliers can constitute the external governance mechanism of enterprises, improve the level of corporate governance, alleviate the agency cost of enterprises, and increase the cash dividends. Specifically, the formation of supplier relationships constitutes a proprietary relationship investment between the buyer and the seller and consequently provides suppliers with access to acquire enterprise proprietary information (19, 20, 48). Accordingly, the supplier can serve as the supervision and certification entity of the enterprise and supervise the opportunistic behavior of the enterprise management, thus reducing the agency cost. In conclusion, the cooperative effect of the supplier relationship alleviates the agency problem while improving the corporate governance environment, and thus it can optimize the enterprise's production and operation activities, generate higher profitability and performance, and increase the company's ability to pay dividends, which is found to be consistent with the conclusion of the "outcome model" of dividend payments of La Porta et al. (42), that is, the company may increase the willingness and the level of the dividend payments.

On the other hand, suppliers can achieve in-depth cooperation by participating in joint research and development of enterprises, offering inventory management for downstream enterprises, or by providing trade discounts and relaxing the term of the commercial credit, which eases the financial constraints of enterprises to some extent, thus enabling enterprises to have more sufficient cash for the dividend payments (40, 49–51). In addition, supplier financing is considered an effective alternative financing instrument that enterprises can use to alleviate financing constraints when the formal system is imperfect. The alleviation of financing constraints implies reducing the cost of external funds and accordingly increasing the cash dividend payments. Based on the above analysis, Hypothesis 1 is proposed.

Hypothesis 1: The supplier relationship is positively correlated with the cash dividend payments.

Supplier Relations and Cash Dividend Policies: Competitive Effect

In the competition, both parties will face the possibility of being misappropriated profit by the party with bargaining advantages. The competitive model of the supplier relationship acts on the corporate cash dividend policy by aggravating agency conflicts and financing constraints.

Due to different risk preferences and return functions faced by the management and shareholders, the manager has a strong advantage to retain more free cash flow based on self-serving behaviors, such as on-the-job consumption, which leads to the appropriation of the cash dividend payments level of the company (33). The subtle relationship between enterprises

and suppliers appropriately provides convenient conditions for the management to retain cash. On the one hand, under the competitive effect, the management takes advantage of its buyer's market advantage and seeks to preserve the free cash flow to a greater extent under the pretext of investing in relational proprietary assets in suppliers, and thus, it provides opportunities to them to increase in-service consumption and personal empire construction and increases agency costs, and thus exacerbating the agency problem and expropriating the cash used for the dividend payments (52). In addition, under the influence of competition, suppliers are more inclined to maximize their own interests and use interests as bait to lobby the management of downstream enterprises to facilitate transactions without considering the real capabilities of suppliers. This leads to an increase in inefficient investment due to a lack of high-quality suppliers. Furthermore, the supplier shows opportunistic behavior toward the enterprise for maximizing its own interests, more specifically, exploiting the profits of downstream enterprises by offering low-quality raw materials, thus resulting in an inefficient relational asset investment, which consequently reduces the value of the company and encroaches on the cash used to pay dividends (53).

The supplier concentration degree formed by the supplier relationship reflects the bargaining advantage of suppliers, which easily leads to the competition effect (54), thus reducing the supply of the commercial credit (55). The reason is that under the circumstance of a powerful position possessed by suppliers, even if they reduce the provision of commercial credit to enterprises, they are less likely to face the risk of enterprises changing suppliers. Under the circumstance that suppliers are highly concentrated and have obvious bargaining advantages, they tend to expropriate the cash flow of enterprises by using their bargaining advantage, for instance, decreasing the supply of the commercial credit, increasing the prices of materials, reducing discounts, elevating cash sales, decreasing credit sales, and shortening the payback periods of goods, and thus increasing the degree of financial constraints of enterprises (55). Consequently, it encroaches on the cash quota of cash dividends issued by enterprises and reduces the tendency and the payment intensity of cash dividends (56). Based on the above analysis, Hypothesis 2 is proposed.

Hypothesis 2: The supplier relationship is negatively correlated with the cash dividend payments.

SAMPLE SELECTION AND DATA

Sample Selection

In this study, the data of Chinese A-share listed companies from the period 2007 to 2020 were selected as initial research samples; all details were taken from the "CSMAR" database and WIND database. In addition, for the supplement of missing values, this study searched its annual report through the website of Oriental Fortune and big data and calculated its missing indicators according to the index calculation formula. In this study, the sample companies with excessive outliers and unavailable values are excluded, whereas for the companies with less unavailable values, the annual reports were searched online, and the unavailable values were calculated according to the index calculation formula. To improve the accuracy of the empirical results, the data were analyzed based on the following steps: in view of the current situation of financial anomalies or continuous losses of ST-class and PT-class listed companies, liquidity constraints were very serious. Hence, the samples of such companies were first excluded. Given that the characteristics of capital expenditure in the financial service industries are vastly different from those of other firms, banks, insurance companies, securities, and other types of companies were excluded, simultaneously, companies with incomplete data on relevant indicators, such as corporate governance, were eliminated. To reduce the estimation error caused by statistical errors or abnormal samples and to consider the influence of extreme values, all data were Winsorize processed at 1% and 99% quantiles.

Model Specification

According to the characteristics of the explained variables and sample data, this study uses the Logit model (1) of panel data to study the influence of the supplier concentration/supplier financing on the cash dividend payment tendency of the listed companies. In cases when some enterprises do not pay cash dividends, the ratio of the cash dividend is 0. This makes the cash dividend payment rate a trailing variable with a lower limit of 0. If the regression method of the ordinary OLS is used, it becomes easy to produce deviation. As a consequence, this study refers to the study of Fama and French (57), Brockman and Unlu (31, 32) to develop the Logit model (1) and the Tobit model (2) of the dividend payments. In models (1) and (2), the subscript i represents the company, t represents the year, the explanatory variable top 5 represents the supplier concentration, credit represents the supplier credit financing, and controls represent a group of control variables. The Logit model and the Tobit model are explained as follows.

Logit model: It takes into consideration the supplier concentration, supplier commercial credit financing, and cash dividend payment tendency:

$$\begin{aligned} \text{Logit}(d_payeri, t) &= \alpha_0 + \alpha_1 Top5/credit_{i,t} + \Sigma\alpha_2 control \\ &+ \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \end{aligned} \quad (1)$$

Tobit model: It considers the supplier concentration, supplier commercial credit financing, and cash dividend payment level:

$$\begin{aligned} \text{Tobit}(divi\text{-}ratio, t) &= \beta_0 + \beta_1 Top5/credit_{i,t} + \Sigma\beta_2 control \\ &+ \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Definition of Variables

Explained Variable

Considering the discrepancy between the cash dividend payment tendency and the cash dividend payment degree, this study refers to the research of Denis and Osobov (58) and uses the cash dividend payment tendency (d_payer) and the payment level

(divi-ratio) to measure the cash dividend policy of the listed companies. Specifically, the cash dividend payment tendency (d_payer) is considered as the dummy variable. If the enterprise pays the cash dividends in the current year, d_payer is considered as 1; otherwise, it is considered as 0. The cash dividend payout level (divi-ratio) is calculated as the ratio of cash dividends to net income.

Explaining Variable

To fully obtain the cooperation and competition status of the supplier relationship strategy, the measurement index of the supplier relationship strategy requires reflecting not only the closeness of the relationship between enterprises and suppliers but also their bargaining advantages. This study refers to the existing literature and applies supplier concentration as one of the indicators to measure the supplier relationship strategy (16, 17, 59, 60). For another dimension of supplier relations, considering the work of Wang and Wang (61), supplier financing that is the net occupation of enterprises, to supplier financing, is selected.

(a) Supplier concentration (Top5): For the measurement of the supplier relationship strategy, most of the earlier studies have used the supplier concentration degree as a proxy. A high proportion of the Top5 implies a close supplier relationship and a high bargaining advantage for suppliers. In this study, the proportion of the purchasing amount of the top five suppliers in the total annual purchasing amount of the listed companies is used as the measure of the supplier concentration.

(b) Supplier commercial credit financing (credit): The existing literature considers that supplier commercial credit financing is the exclusive asset investment existing in the transaction of the supplier relationship strategy, which suggests the higher the proportion of enterprises' commercial credit financing with suppliers, the closer the relationship with suppliers and the stronger the bargaining advantage gained by suppliers. This study refers to the measurement method of the utilization level of supplier financing proposed by Wang and Wang (61), which is measured as the proportion of the enterprise's net occupation of suppliers to the cost of sales. The larger the proportion of the enterprise's net occupation of suppliers, the more the supplier financing, and hence the indicator is calculated as (accounts payable + notes payable – accounts prepaid)/cost of sales.

Other Control Variables

In addition, the model also controls other factors that affect dividend payments. Refer to the studies of Chae et al. (33) and Brockman and Unlu (31), Profit, growth, ROE, age, and state are selected as control variables in this study, and the calculation method is as follows: Profit, that is, the ratio of net profit to total assets. The growth of the company is expressed by the growth rate of the main business income. ROE refers to return on equity, and the company's listing age is the same year —- the company's listing year +1, and the property rights of the company (State). In addition, the control variables also included the industry dummy variable Ind and the annual dummy variable Year.

TABLE 1 | Descriptive statistics.

VarName	Obs	Mean	SD	Min	Median	Max
dpayer	34,237	0.711	0.453	0.000	1.000	1.000
divi-ratio	35,005	0.275	1.033	−64.428	0.204	107.407
divi-ratio2	34,459	0.116	0.357	0.000	0.014	2.623
Age	34,459	10.367	7.058	1.000	9.000	27.000
Growth	35,005	0.152	0.338	−0.591	0.108	1.844
Profit	35,005	3.999	6.322	−26.552	3.920	21.131
ROE	35,005	0.077	0.128	−0.573	0.079	0.419
State	35,005	0.397	0.489	0.000	0.000	1.000
Top 5	35,005	0.243	0.235	0.000	0.209	0.943
Credit	35,005	0.263	0.282	−0.479	0.209	1.461

EMPIRICAL EVIDENCE

Descriptive Statistics

Table 1 presents the descriptive statistical characteristics of the main variables. From the descriptive statistics, it can be observed that the mean value of d_payer for the dividend payments is 0.711, which indicates that more than 70% of the sample listed companies have paid cash dividends. The mean value of supplier concentration is 24.3%, which indicates that China's listed companies show a relatively low degree of supplier concentration. In addition, the average ratio of enterprise supplier financing to the cost of sales (credit) is measured as 26.3%, with a maximum value of 146.1% and a minimum value of −47.9%. It can be concluded that Chinese A-share listed companies have great differences in the capital occupation of suppliers.

Basic Regression Results of the Supplier Relationship Strategy and the Enterprise Cash Dividends

Table 2 presents the regression results of the main effects of models (1) and (2), that is, the regression result of the relationship between supplier concentration/supplier financing, dividend payout tendency, and dividend payout rate. Columns (1) and (2) of the table represent the relationship between the supplier concentration(top5)/supplier financing credit(credit) and the cash dividend payment tendency, respectively. As can be observed from column (1), the coefficient of the supplier concentration Top5 and the dividend payment tendency is −0.749, which is significant at the level of 1%, indicating that the supplier concentration degree is negatively correlated with the dividend payment tendency. From the results of column (2), it can be concluded that the correlation between the supplier financing credit and the dividend payments preference d_payer is negative. Columns (4) and (5) of the table show the Tobit regression results of the model (2), namely, the relation between supplier concentration top5/supplier financing credit and the cash dividend payment level. As can be observed from column (4), the coefficient between the supplier concentration and the dividend payment level is −0.065, which is significant at the level of 1%. From column (5) of the table, it can be concluded that the coefficient of supplier credit financing and dividend

TABLE 2 | Supplier relationship strategy top5/credit and dividend payout propensity (logit), dividend payout level (Tobit) regression results.

	(1) d_payer	(2) d_payer	(3) d_payer	(4) divi-ratio	(5) divi-ratio	(6) divi-ratio
Age	−0.092***	−0.091***	−0.093***	−0.011***	−0.011***	−0.012***
	(−36.05)	(−35.61)	(−36.22)	(−30.34)	(−30.13)	(−30.54)
Growth	−0.113**	−0.131***	−0.118***	−0.097***	−0.099***	−0.098***
	(−2.51)	(−2.92)	(−2.62)	(−13.34)	(−13.62)	(−13.49)
Profit	0.204***	0.197***	0.200***	0.019***	0.018***	0.018***
	(29.63)	(28.84)	(29.07)	(24.45)	(23.08)	(23.21)
ROE	4.487***	4.721***	4.577***	0.324***	0.355***	0.344***
	(14.69)	(15.49)	(15.01)	(8.49)	(9.25)	(8.97)
State	0.396***	0.416***	0.399***	0.021***	0.023***	0.022***
	(11.49)	(12.10)	(11.59)	(3.87)	(4.30)	(3.98)
Top5	−0.749***		−0.776***	−0.065***		−0.070***
	(−10.51)		(−10.85)	(−5.84)		(−6.32)
Credit		−0.353***	−0.385***		−0.074***	−0.077***
		(−6.24)	(−6.79)		(−8.22)	(−8.57)
_cons	−0.835***	−0.817***	−0.843***	0.081**	0.082**	0.081**
	(−4.25)	(−4.16)	(−4.30)	(2.37)	(2.39)	(2.37)
ind	Yes	Yes	Yes	Yes	Yes	Yes
yr	Yes	Yes	Yes	Yes	Yes	Yes
N	34,237	34,237	34,237	34,459	34,459	34,459

**is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).*

payment intensity is significantly negative at the level of 1%, with a coefficient of −0.074. Therefore, the supplier concentration degree/supplier financing is negatively correlated with the cash dividend payment level, which proves Hypothesis 2.

Robustness Test

The above empirical results may be affected by quantitative problems, such as inverse causality problems, missing variables, and sample selection bias. Therefore, this study performed the following robustness test to ensure the reliability of the above conclusions.

Reverse Causality Problem

The selection of the dividend policy may affect the enterprise's supplier concentration and supplier financing. In other words, there is a potential reverse causality between the supplier relationship strategy and the cash dividend payment level. To test this problem, this study considers the dividend payment tendency of the next phase $dpayer_{t+1}$ and the cash dividend payment level of the next phase $divi\text{-}ratio_{t+1}$ as the explained variables. The regression results are presented in **Table 3**. The results in column (1) prove that there is a significant negative correlation between the supplier concentration (top5) and the cash dividend payment level of the next phase at the level of 1%. In addition, based on the results presented in column (3), the supplier concentration is negatively correlated with the dividend payment tendency of the next phase, which is found to be consistent with the above conclusion. Similarly, based on the results presented in column (2), it is proved that supplier commercial credit financing is significantly negatively correlated with the cash dividend payment level of the next phase, which is found to be consistent with the above conclusion. The results of column (4) also prove the robustness of the results.

Replace Key Variables

To ensure the robustness of the results, this study refers to Kao and Chen (38) and Xu and Xu (62) to replace the measurement index of dividend payment level. As a consequence, this study uses the cash dividend yield divi-ratio2, which is calculated as the ratio of the cumulative cash dividend paid to the closing price of the year substituted for the dividend payment level. The results are presented in columns (5) and (6) of **Table 3**. Column (5) shows the regression results of the relationship between the supplier concentration top5/supplier commercial credit financing and the original variable divi-ratio. Column (6) shows a significant relationship between top5, credit, and the substitution variable divi-ratio2, indicating that the result is still negatively correlated at the level of 1%, which is found to be consistent with the previous conclusion, thus proving the robustness of the results.

Propensity Score Matching (PSM)

Since the model may exist as the endogeneity of missing variables, it may be related to both the dividend payments and the supplier relationship strategy. Specifically, if the control variables in the model fail to well capture the discrepancies between the centralized and decentralized suppliers, as well as the diversities in the characteristics of the supplier financing

TABLE 3 | Inverse causality problem test and the substitution key variable method.

	Inverse causality problem test				Substitution key variable method	
	(1) divi-ratiot+1	(2) divi-ratiot+1	(3) dpayert+1	(4) dpayert+1	(5) divi-ratio	(6) divi-ratio2
Age	−0.010***	−0.010***	−0.076***	−0.075***	−0.012***	−0.002***
	(−24.85)	(−24.64)	(−29.65)	(−29.31)	(−30.54)	(−6.54)
Growth	−0.087***	−0.089***	−0.084*	−0.099**	−0.098***	−0.052***
	(−11.05)	(−11.30)	(−1.90)	(−2.24)	(−13.49)	(−7.86)
profit	0.018***	0.017***	0.131***	0.126***	0.018***	0.008***
	(21.65)	(20.43)	(22.30)	(21.50)	(23.21)	(11.16)
ROE	0.318***	0.346***	3.915***	4.075***	0.344***	1.072***
	(7.78)	(8.44)	(14.37)	(14.93)	(8.97)	(28.00)
State	0.020***	0.023***	0.354***	0.370***	0.022***	0.073***
	(3.50)	(3.88)	(10.41)	(10.89)	(3.98)	(14.79)
Top5	−0.060***		−0.577***		−0.070***	−0.223***
	(−4.96)		(−8.19)		(−6.32)	(−21.96)
Credit		−0.075***		−0.353***	−0.077***	−0.063***
		(−7.62)		(−6.19)	(−8.57)	(−7.70)
_cons	−0.021	−0.019	−0.831***	−0.815***	0.081**	−0.319***
	(−0.58)	(−0.53)	(−4.32)	(−4.24)	(2.37)	(−10.05)
ind	Yes	Yes	Yes	Yes	Yes	Yes
yr	Yes	Yes	Yes	Yes	Yes	Yes
N	30,629	30,629	30,409	30,409	34,459	34,459

**is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).*

scale, the measurement index of the supplier relationship strategy will induce a nonlinear effect. Due to the special types of the Logit and Tobit models, this study uses the propensity score matching method to control the endogeneity of missing variables. This study refers to the research methods proposed by Jiao and Zhang (30) and Meng and Bai (63). The treatment and control groups were divided considering the fact whether the quantile of the Top5 of the enterprise was >60% of the sample (60% quantile of the Top5 and credit variables were used to keep the result unchanged). The propensity score was calculated by using Logit regression. The asset size LNsize, financial leverage Lev, sales revenue Growth rate, property right property State, company Age, market value to the book value ratio Mkt, and industry HHI were used as matching variables. Subsequently, the propensity score matching analysis was performed based on the data of the enterprise in the current year. The model used the most common method of "nearest-neighbor matching" to match the treatment and the control groups. The results are presented in **Table 4**. In column (13) of the table, the *P* values are all >0.05. After matching all variables, no significant difference was observed between the treatment and the control groups, indicating that there is no statistical difference between them. Moreover, the mean difference between the groups is not found to be significant, which basically satisfies the balance test of PSM. In this study, after the PSM test, the results are presented in **Table 5**. Comparing this result with the main effect result presented in **Table 2**, it is found that in column (1) of **Table 5**, for the Tobit model of the dividend payment level, the regression coefficient of Top5 is −0.072. The regression coefficients of the top5 were found to be significantly negatively correlated, which proves the robustness of the results. Similarly, the regression coefficients of credit in column (1) of **Table 5** demonstrate the robustness of the results.

INTERMEDIARY MECHANISM TEST OF THE SUPPLIER RELATIONSHIP STRATEGY AND THE CASH DIVIDEND PAYMENT LEVEL

The intermediary mechanism model can analyze the process and mechanism of the influence of independent variables on dependent variables. This study argues that supplier concentration and supplier financing act on corporate cash dividends through agency costs and financing constraints. The most popular causal steps approach for testing is the Baron and Kenny (64) causal steps approach and the Sobel (65) approach. Therefore, by combining the advantages of the above methods, this study uses the mediating effect analysis method summarized by Baron and Kenny (64) and Wen and Ye (66) to examine the influence mechanism of supplier relationship strategy on cash dividends.

TABLE 4 | Average processing effect of PSM (ATT).

(1) Variable	(2) Sample	(3) Treated	(4) Controls	(5) Difference	(6) S.E	(7) T-stat
divi-ratio	Unmatched	0.273659	0.27030	0. 00335	0 0.0095	0.35
	ATT	0. 273699	0. 29912	−0. 02542	0.0158	−1.60**
Variable	Sample	Treated	Controls	Difference	S.E	T-stat
d_payer	Unmatched	0.7157	0.7048	0.0109	0 0.0050	2.18**
	ATT	0.7159	0.7244	−0.0085	0. 0068	−1.24*

(8) Variable	(9) Treated	(10) control	(11) %bias	(12) *t*	(13) *p*>|t|	(14) V (T)/V (C)
Age	10.031	10.128	−1.4	−1.11	0.269	1.23*
Growth	0.1510	0.1473	1.1	0.89	0.371	1.29*
Lev	0.3984	0.4067	−0.2	−1.9	0.057	0.72*
Mkt	2.618	2.816	−1.1	−0.8	0.423	0.02*
State	0.3233	0.3152	1.7	1.44	0.150	
HHI	0.1066	0.10638	0.2	0.21	0.833	1.09*
LNSize	21.838	21.858	1.9	1.66	0.097	0.96*

*is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).

TABLE 5 | Propensity score matching (PSM processing).

	(1) divi-ratio	(2) d_payer
Age	−0.011***	−0.092***
	(−30.14)	(−35.89)
Growth	−0.099***	−0.114**
	(−13.55)	(−2.53)
profit	0.018***	0.202***
	(22.59)	(29.05)
ROE	0.338***	4.534***
	(8.85)	(14.76)
State	0.019***	0.396***
	(3.50)	(11.44)
Top5	−0.072***	−0.782***
	(−6.49)	(−10.90)
Credit	−0.077***	−0.382***
	(−8.55)	(−6.69)
_cons	0.087**	−0.862***
	(2.53)	(−4.39)
ind	Yes	Yes
yr	Yes	Yes
N	34,070	34,070

*is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).

Mediation Variables and Model Specification

Mediation variables: The agent cost AC and financing constraint FC are considered as mediation variables. This study refers to the existing literature. For the measurement of the first type of agency cost, this study refers to the approaches proposed by Singh and Davidson III (36), Li (67) and Jiang et al. (68). In this study, the management expense ratio AC is used to measure the first type of agency cost, and the calculation method is measured as the ratio of the management expense to the main business income. This study exploits the SA index developed by Hadlock and Pierce (69) to measure the degree of the financing constraint (FC) of the listed companies. SA $= -0.737 \times$ company size $+ 0.043$ * square of the total assets $-0.04 \times$ company listing time. The larger the SA index, the smaller the degree of the financing constraint. To test the effect mechanism of the supplier relationship strategy on the corporate cash dividends, this study uses the mediating effect analysis method proposed by Baron and Wen, and this specific model is expressed as follows:

$$AC = \chi_0 + \chi_1 Top5/credit_{i,t} + +\Sigma\chi_2 control + \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \quad (3a)$$

$$FC = \gamma_0 + \gamma_1 Top5/credit_{i,t} + +\Sigma\gamma_2 control + \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \quad (3b)$$

$$\text{Tobit(d}ivi - ratio_{i,t})/\log it(d_\text{payer}) = \varphi_0 + \varphi_1 Top5/credit_{i,t} + \varphi_2 AC + \sum \varphi_3 controls + \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \quad (4a)$$

$$\text{Tobit(d}ivi - ratio_{i,t})/\log it(d_payer) = \delta_0 + \delta_1 Top5/credit_{i,t} + \delta_2 FC + \sum \delta_3 controls + \Sigma Year + \Sigma Ind + \varepsilon_{i,t} \quad (4b)$$

Empirical Results of the Mechanism Test

Table 6 presents the results of the mechanism test for the relationship between the supplier concentration and the cash

TABLE 6 | The mechanism test of supplier concentration acts on cash dividend payment tendency and dividend payment level.

	(1) AC	(2) divi-ratio	(3) d_payer	(4) fc	(5) divi-ratio	(6) d_payer
Age	0.001***	−0.013***	−0.095***	−0.004**	−0.013***	−0.014***
	(3.04)	(−17.70)	(−19.51)	(−2.29)	(−17.70)	(−22.21)
Growth	−0.025***	−0.108***	−0.398***	0.101***	−0.106***	−0.055***
	(−5.47)	(−11.03)	(−6.99)	(5.07)	(−10.85)	(−6.74)
Profit	0.005***	0.022***	0.340***	−0.040***	0.025***	0.032***
	(5.82)	(12.25)	(19.92)	(−9.49)	(13.62)	(22.01)
ROE	−0.402***	0.070	0.691	2.948***	−0.065	0.251***
	(−10.57)	(0.79)	(0.97)	(13.97)	(−0.72)	(3.64)
State	0.018***	−0.033***	−0.324***	−0.377***	−0.019**	−0.013
	(3.79)	(−3.50)	(−5.32)	(−19.38)	(−2.00)	(−1.58)
Top5	0.023***	−0.042***	−0.573***	−0.600***	−0.016	−0.038***
	(3.54)	(−3.15)	(−6.82)	(−21.39)	(−1.19)	(−3.19)
AC		−0.446***	−3.507***			
		(−7.69)	(−9.06)			
FC					0.055***	0.085***
					(11.38)	(20.27)
_cons	0.085***	0.323***	2.296***	−0.095	0.243***	0.802***
	(3.15)	(6.35)	(6.91)	(−0.93)	(4.79)	(17.97)
ind	Yes	Yes	Yes	Yes	Yes	Yes
yr	Yes	Yes	Yes	Yes	Yes	Yes

*is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).

dividend payment preference d_payer as well as the cash dividend payment level divi-ratio, respectively. The results show that the supplier concentration affects the willingness and the level of the cash dividend payments by increasing the agency cost and the financing constraint. The specific testing steps followed in the intermediary effect mechanism are the following.

Step 1: Implement the regression test for the main effect, namely, models (1) and (2). The regression results are presented in **Table 2**, and the coefficient is found to be significantly negative, which indicates that the supplier concentration is significantly negatively correlated with the cash dividend payment tendency and the cash dividend payment level of enterprises.

Step 2: Regression is performed on models (3a) and (3b), that is, the intermediary variables (agency cost AC and financing constraint FC) were respectively regression with the supplier concentration (top5) of the explanatory variable. The results are presented in columns (1) and (4) of **Table 6**. As can be observed from the table, the higher the concentration of suppliers, the higher the agency cost of the enterprise would be (the coefficient between top5 and AC is measured as 0.023, which is significantly positive). At the same time, the high concentration of suppliers will increase the degree of the financing constraint of enterprises (the coefficient of top5 and FC is measured as −0.600, which is significantly negative, and the dependent variable FC represents the degree of the financing constraint; the higher the FC, the lower the degree of the financing constraint). Therefore, top5 is found to be positively correlated with the degree of financing constraint.

Step 3: For models (4a) and (4b), the explained variable (divi-ratio and d_payer), the explanatory variable (supplier concentration (top5)), and the intermediary variable (agent cost AC and financing constraint FC) were used in the same model for regression to test the significance and the positive or the negative value of the coefficient. The agent cost AC is considered as the first intermediary variable, whose coefficients with the cash dividend payment level and the payment tendency are calculated as −0.446 and −3.507 in columns (2) and (3), respectively. The coefficient is significantly negative, which indicates that the agency cost is negatively correlated with the cash dividends and the cash dividend payment tendency. This finding is consistent with the agency cost theory of dividends. Similarly, for the financing constraint of the second intermediary variable, FC, whose coefficients with the cash dividend payment level and the payment tendency are 0.055 and 0.085 in columns (5) and (6), the significance is positive. Since FC represents the degree of the financing constraint, the lower the FC index, the higher the degree of the enterprise financing constraint would be. Therefore, it can be assumed that the financing constraint is negatively correlated with the cash dividend payment level and the payment tendency, which is found to be consistent with the financing constraint theory of dividends, that is, enterprises with a higher financing constraint level show a lower cash dividend payment level.

Step 4: It presents the validation of the mediating effect according to the significance of the coefficient and its symbol. When the dividend payment level (divi-ratio) and the dividend payment preference (d_payer) are considered as the explained

TABLE 7 | Supplier financing and cash dividend payment level mechanism test.

	(1) AC	(2) divi-ratio	(3) d_payer	(4) fc	(5) divi-ratio	(6) d_payer
Age	0.001***	−0.013***	−0.095***	−0.003*	−0.013***	−0.094***
	(3.24)	(−17.76)	(−19.48)	(−1.95)	(−17.84)	(−18.82)
Growth	−0.024***	−0.109***	−0.422***	0.081***	−0.107***	−0.457***
	(−5.28)	(−11.21)	(−7.43)	(3.94)	(−10.98)	(−7.87)
Profit	0.006***	0.021***	0.329***	−0.041***	0.024***	0.374***
	(6.85)	(11.43)	(19.09)	(−9.11)	(12.82)	(21.18)
ROE	−0.428***	0.126	1.149	3.144***	−0.029	−0.677
	(−11.27)	(1.41)	(1.59)	(14.14)	(−0.32)	(−0.95)
State	0.020***	−0.035***	−0.343***	−0.395***	−0.020**	−0.157**
	(4.22)	(−3.76)	(−5.63)	(−19.41)	(−2.13)	(−2.50)
Credit	0.068***	−0.072***	−0.303***	0.108***	−0.092***	−0.569***
	(8.25)	(−4.21)	(−2.82)	(3.10)	(−5.43)	(−5.22)
AC		−0.428***	−3.453***			
		(−7.36)	(−8.91)			
FC					0.056***	0.731***
					(11.98)	(20.86)
_cons	0.072***	0.337***	2.171***	−0.401***	0.281***	1.681***
	(2.67)	(6.59)	(6.56)	(−3.99)	(5.52)	(4.87)
ind	Yes	Yes	Yes	Yes	Yes	Yes
yr	Yes	Yes	Yes	Yes	Yes	Yes

*is significant at 0.1 level, **is significant at 0.05 level, ***is significant at 0.01 level (the same below).

variables, based on the judgmental criteria of the intermediary effect, the agent cost AC is selected as an example. According to the results that the coefficients β_1, α_1, χ_1, φ_1, and φ_2 are all significant, top5 is found to be significantly negatively correlated with the dividend payment level and the dividend payment tendency, as presented in columns (2) and (3) of **Table 6**, which shows the partial intermediary effect of the agency cost on the relationship between the supplier concentration and the cash dividends. The intermediary effect of the agency cost and the financing constraint on the relationship between the supplier concentration and the cash dividends is explained as follows: combining the results of models (1), (2), (3a), and (4a), the following results can be obtained: the coefficient of the explanatory variable top5 in model (3a) is measured as 0.023, and its product with the coefficient of the agent cost AC in model (4a) is found to be negative, i.e., −0.446, which is the same as the negative sign of the coefficient of the explanatory variable top5 in model (4a), i.e., −0.042. This result indicates that part of the intermediary effect of the agent cost exists, and the total proportion of the intermediary effect is determined as 0.023 * (−0.446)/(−0.042) = −0.24. Similarly, the partial mediating effect of the agency cost on the relationship between top5 and dividend payment tendency can also be proved. The proportion of the agency cost on the mediating effect of the dividend payment tendency is 0.023 * (−3.507)/ −0.573 = 0.14. Similarly, the proportion of the mediating effect of the financing constraint to the total effect of the cash dividend payment level is determined as −0.600 * 0.055/(−0.016) = 2.0625, whereas the total effect of the cash dividend payout tendency is determined as −0.600*0.085/−0.038 = 1.34. To sum up, the increase in supplier concentration leads to an increase in agency costs and financing constraints of enterprises, thus reducing cash dividends.

Similarly, **Table 7** presents the regression results of the mechanism of the supplier commercial credit and cash dividend payment dynamics. In **Table 7**, the supplier commercial credit (credit) is considered as the explanatory variable, and the dividend payment level (divi-ratio) is considered as the explained variable. The steps of the intermediary effect test of the supplier commercial credit financing and cash dividend payment level are also the same as discussed in the previous section, and the corresponding results are presented in **Table 7**. Based on the principle of the intermediary effect test, the mediating effect of agency cost on supplier financing, incentive payment intensity, and dividend payment tendency is 0.404 and 0.77, respectively. In particular, because the sign symbols are different, financing constraint has suppressing effects on supplier financing and dividend payment level and dividend payment tendency, which are 0.065 and 0.138, respectively.

To sum up, the supplier concentration degree/supplier financing primarily affects the cash dividend payment tendency and the cash dividend payment level by increasing the agency cost and the financing constraint degree. To more clearly present the relationship between supplier relationship strategy and cash dividends as well as the intermediary mechanism, the relationship between the above variables is sorted out in **Figure 1**, and the mediating effect of each component is shown in **Table 8**.

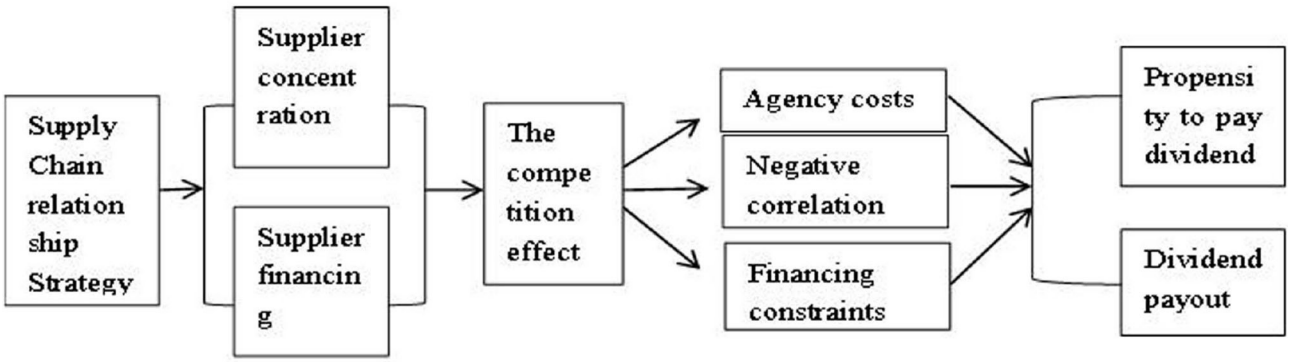


FIGURE 1 | Relationship and mechanism of supplier relationship strategy and cash dividend.

TABLE 8 | Results of intermediary effect between supplier relationship strategy and cash dividend: Intermediary effect of agency cost and financing constraint.

Intermediate effect	Cash dividend payment level				Cash dividend payout tendency			
	Agency cost	Type of intermediate effect	Financing constraints	Type of intermediate effect	Agency cost	Type of intermediate effect	Financing constraints	Type of intermediate effect
Supplier concentration	0.24	Partial intermediate effect	2.0625	Complete mediation effect	0.14	Partial intermediate effect	1.34	Partial intermediate effect
Supplier financing	0.404	Partial intermediate effect	0.065	Suppressing effect	0.77	Partial intermediate effect	0.138	Suppressing effect

CONCLUSION

The relationship transaction is increasingly affecting corporate governance and financial management. Supplier relationship strategy, i.e., choosing a centralized supplier strategy or decentralized supplier strategy and how to leverage the right supplier financing scale, has theoretical and practical significance on whether and how to influence corporate cash dividends. More importantly, the characteristic of the supplier relationship strategy is that the appearance of cooperation and competition alternately results in the various elusive economic consequences or they coexist. Therefore, enterprises need to timely judge the different economic consequences under the state of dynamic supplier relationship based on big data and timely adjust the operation and financial decisions. This study explores how the cooperative effect and competition effect produced by supplier relationship affect corporate cash dividends. It is demonstrated that supplier concentration and supplier financing, are negatively correlated with the cash dividends. Consequently, the competitive effect of the supplier relationship is proved. In the test of the intermediary mechanism of the influence of the supplier relationship strategy on the cash dividends, it is observed that the competitive effect of the supplier relationship strategy aggravates the agency conflict and the financing constraint of enterprises, and thus shows a negative effect on the willingness and level of the cash dividend payments. Based on Chinese data analysis, this study provides a decision-making basis for enterprises to maximize the positive effects of supply chain relationship strategy, and provides information basis for regulators and external investors to make decisions at the supply chain level.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LS contributed to methodology and writing of the article, software, data, and writing-original draft. XY contributed to reviewing and editing the article. Both authors contributed to the article and approved the submitted version.

FUNDING

XY acknowledges the National Natural Science Foundation of China (No. 72062027).

REFERENCES

1. Matopoulos A, Didonet S, Tsanasidis V, Fearne A. The role of perceived justice in buyer-supplier relationships in times of economic crisis. *J Purch. Supply Manag.* (2019) 25:100554. doi: 10.1016/j.pursup.2019.100554
2. Baraldi E, Ratajczak-Mrozek M. From supplier to center of excellence and beyond: the network position development of a business unit within "Ikea Industry". *J Bus Res.* (2019) 100:1–15. doi: 10.1016/j.jbusres.2019.03.008
3. Chen IJ, Paulraj A, Lado AA. Strategic purchasing, supply management, and firm performance. *J Oper Manag.* (2004) 22:505–23. doi: 10.1016/j.jom.2004.06.002
4. Amoako-Gyampah K, Boakye KG, Adaku E, Famiyeh S. Supplier Relationship management and firm performance in developing economies: a moderated mediation analysis of flexibility capability and ownership structure. *Int J Prod Econ.* (2019) 208:160–70. doi: 10.1016/j.ijpe.2018.11.021
5. Das A, Narasimhan R, Talluri S. Supplier integration—finding an optimal configuration. *J Oper Manag.* (2006) 24:563–82. doi: 10.1016/j.jom.2005.09.003
6. Bankvall L, Dubois A, Lind F. Conceptualizing business models in industrial networks. *Ind Mark Manag.* (2017) 60:196–203. doi: 10.1016/j.indmarman.2016.04.006
7. Frohlich MT. E-Integration in the Supply Chain: barriers and Performance. *Decis Sci.* (2002) 33:537–56. doi: 10.1111/j.1540-5915.2002.tb01655.x
8. Corsten D, Kumar N. Do Suppliers Benefit from Collaborative Relationships with Large Retailers? An empirical investigation of efficient consumer response adoption. *J Mark.* (2005) 69:80–94. doi: 10.1509/jmkg.69.3.80.66360
9. Alletto A, Bruccoleri M, Mazzola E, Ramanathan U. Collaboration Experience in the supply chain of knowledge and patent development. *Prod Plan Control.* (2017) 28:574–86. doi: 10.1080/09537287.2017.1309712
10. Dehning B, Richardson VJ, Zmud RW. The Financial performance effects of it-based supply Chain management systems in manufacturing firms. *J Oper Manag.* (2007) 25:806–24. doi: 10.1016/j.jom.2006.09.001
11. Wang X. Contract motivation, information sharing and dynamic coordination of supply Chain. *Manag World*. (2005) 4:106–15.
12. Song J. Application of Ahp model in supply chain selection. *J Hubei Univ Econ*. (2010) 7:67–69.
13. Zaremba BW, Bode C, Wagner SM. New venture partnering capability: an empirical investigation into how buying firms effectively leverage the potential of innovative new ventures. *J Supply Chain Manag.* (2017) 53:41–64. doi: 10.1111/jscm.12116
14. Dass N, Kale JR, Nanda V. Trade credit, relationship-specific investment, and product market power. *Rev Finance.* (2015) 19:1867–923. doi: 10.1093/rof/rfu038
15. Barrot JN. Trade credit and industry dynamics: evidence from trucking firms. *J Finance.* (2016) 71:1975–2016. doi: 10.1111/jofi.12371
16. Sun L, Zhai S,Wang Z. Supplier relationship, social trust and financing effect of commercial credit. *Soft Sci.* (2017) 31:71–4. doi: 10.13956/j.ss.1001-8409.2017.02.15
17. Wang Y, Liu Z. Supplier relations and corporate cash holdings: Empirical evidence from listed manufacturing companies in China. *Audit Econ Res.* (2016) 1:83–91. doi: 10.3969/j.issn.1004-4833.2016.01.009
18. Bae KH, Kang JK, Wang J. Employee treatment and firm leverage: a test of the stakeholder theory of capital structure. *J financ econ.* (2011) 100:130–53. doi: 10.1016/j.jfineco.2010.10.019
19. Banerjee S, Dasgupta S, Kim Y. Buyer-supplier relationships and trade credit. *HKUST Working Paper*. (2004) 1–49. doi: 10.2139/ssrn.590482
20. Kale JR, Shahrur H. Corporate capital structure and the characteristics of suppliers and customers. *J financ econ.* (2007) 83:321–65. doi: 10.1016/j.jfineco.2005.12.007
21. Zhao L, Sun LY, Li G. Research on the relationship between supply chain integration and enterprise performance in Chinese Manufacturing Enterprises. *J Manag Eng.* (2011) 25:1–9. doi: 10.3969/j.issn.1004-6062.2011.03.001
22. Choi TY, Krause DR. The supply base and its complexity: implications for transaction costs, risks, responsiveness, and innovation. *J Oper Manag.* (2006) 24:637–52. doi: 10.1016/j.jom.2005.07.002
23. Cen L, Dasgupta S, Sen R. Discipline or disruption? Stakeholder relationships and the effect of takeover threat management. *Science.* (2016) 62:2820–41. doi: 10.1287/mnsc.2015.2252
24. Johnson WC, Kang JK, Yi S. The Certification role of large customers in the new issues market. *Financ Manag.* (2010) 39:1425–74. doi: 10.1111/j.1755-053X.2010.01118.x
25. Intintoli VJ, Serfling MA, Shaikh S. The negative spillover effects of ceo turnovers: Evidence from firm-supplier relations. *SSRN Electr J.* (2012). doi: 10.2139/ssrn.2178148
26. Kale J, Kedia S, Williams R. *Product Market Linkages and Managerial Risk Taking*. Georgia State University. Working Paper. (2013). p. 1–49. Available online at: http://en-recanati.tau.ac.il/sites/nihul_en.tau.ac.il/files/media_server/Recanati/management/seminars/account/Kale.pdf
27. Raman K, Shahrur H. Relationship-specific investments and earnings management: evidence on corporate suppliers and customers. *Acc Rev.* (2008) 83:1041–81. doi: 10.2308/accr.2008.83.4.1041
28. Wang X, Fan G. *Report of China's Market Index by Province*. (2016). Beijing: Social Sciences Academic Press. (2017).
29. Wang J. Do Firms' Relationships with Principal Customers/Suppliers Affect Shareholders' Income? *J Corp Finance.* (2012) 18:860–78. doi: 10.1016/j.jcorpfin.2012.06.007
30. Jiao X, Zhang P. Does customer concentration affect dividend policy: governance effect or risk effect. *J Guangdong Univ Finance Econ.* (2017) 4:70–80.
31. Brockman P, Unlu E. Dividend policy, creditor rights, and the agency costs of debt. *J Financ Econ.* (2009) 92:276–99. doi: 10.1016/j.jfineco.2008.03.007
32. Brockman P, Unlu E. Earned/contributed capital, dividend policy, and disclosure quality: an international study. *J Bank Finance.* (2011) 35:1610–25. doi: 10.1016/j.jbankfin.2010.11.014
33. Chae J, Kim S, Lee EJ. How corporate governance affects payout policy under agency problems and external financing constraints. *J Bank Finance.* (2009) 33:2093–101. doi: 10.1016/j.jbankfin.2009.05.003
34. Lu X, Wang J. Research on the correlation between corporate governance and cash dividend: a dual perspective based on agency cost and financing constraint. *Financ Rev*. (2011) 5:91–8.
35. Feng Y, Zhu J, Siu TK. Optimal risk exposure and dividend payout policies under model uncertainty. *Insur Math Econ.* (2021) 100:1–29. doi: 10.1016/j.insmatheco.2021.03.029
36. Singh M, Davidson III WN. Agency costs, ownership structure and corporate governance mechanisms. *J Bank Finance.* (2003) 27:793–816. doi: 10.1016/S0378-4266(01)00260-6
37. Tahir M, Ibrahim H, Zulkafli AH, Mushtaq M. Corruption, national culture, law and dividend repatriation policy. *J Multinatl Financ Manag.* (2020) 57:100658. doi: 10.1016/j.mulfin.2020.100658
38. Kao L, Chen A. How product market competition affects dividend payments in a weak investor protection economy: evidence from Taiwan. *Pac-Basin Finance J.* (2013) 25:21–39. doi: 10.1016/j.pacfin.2013.08.004
39. Qin W, Liang Q, Jiao Y, Lu M, Shan Y. Social trust and dividend payouts: evidence from China. *Pac-Basin Finance J.* (2022) 72:101726. doi: 10.1016/j.pacfin.2022.101726
40. Qu W, Xie Y, Ye Y. Information asymmetry, financing constraint and investment-cash flow sensitivity: an empirical study based on market microstructure theory. *Econ Res*. (2011) 6:105–17. Available online at: http://dspace.xmu.edu.cn:8080/dspace/handle/2288/17371
41. Jensen MC. Agency costs of free cash flow, corporate finance, and takeovers. *Am Econ Rev.* (1986) 76:323–9.
42. La Porta R. Lopez-de-Silanes F, Shleifer A, Vishny RW. Agency Problems and Dividend Policies around the World. *J Finance.* (2000) 55:1–33. doi: 10.1111/0022-1082.00199
43. Luo Q, Xiao W, Xia X. Financing constraints or overinvestment—evidence on investment-cash flow sensitivities of Chinese listed companies. *China Ind Econ*. (2007) 103–10.
44. Yu L, Liang T. Governance effect of dividend policy – from the perspective of balancing financing constraints and agency costs, second issue. *Soft Sci*. (2013) 27:67–70. doi: 10.3969/j.issn.1001-8409.2013.02.014
45. Rozeff MS. Growth, beta and agency costs as determinants of dividend payout ratios. *J Financ Res.* (1982) 5:249–59. doi: 10.1111/j.1475-6803.1982.tb00299.x

46. Zhong Y, Lu X. Financing constraint, agency cost and information content of cash dividend policy. *Soft Sci*. (2013) 27:140–4. doi: 10.3969/j.issn.1001-8409.2013.12.030
47. Li H, Li X, Sun J. A review and analysis on the research status of supply chain partnership management. *J Manag Eng*. (2008) 22:148–51. doi: 10.3969/j.issn.1004-6062.2008.02.028
48. Titman S. The effect of capital structure on a firm's liquidation decision. *J Financ Econ*. (1984) 13:137–51. doi: 10.1016/0304-405X(84)90035-7
49. Whited TM, Wu G. Financial constraints risk. *Rev Financ Stud.* (2006) 19:531–59. doi: 10.1093/rfs/hhj012
50. Gugler K. Corporate governance, dividend payout policy, and the interrelation between dividends, R&D, and capital investment. *J Bank Finance.* (2003) 27:1297–321. doi: 10.1016/S0378-4266(02)00258-3
51. Lian Y, Zhong F, A. Study on the dynamic adjustment mechanism of capital structure of China's listed companies. *Southern Econ*. (2007) 23–38. doi: 10.3969/j.issn.1000-6249.2007.01.003
52. Shin H-H, Stulz RM. Are internal capital markets efficient? *Q J Econ*. (1998) 113:531–52. doi: 10.1162/003355398555676
53. Gelderman CJ, Semeijn J, Verhappen M. Buyer opportunism in strategic supplier relationships: triggers, manifestations and consequences. *J Purch Supply Manag.* (2019) 26:100581. doi: 10.1016/j.pursup.2019.100581
54. Dowlatshahi S. Bargaining power in buyer-supplier relationships. *Prod Inventory Manag J*. (1999) 40:27.
55. Ma L, Zhang M, Yi Z. Does the relationship between suppliers and customers affect the business credit of enterprises? Based on the empirical test of chinese listed companies. *Econ Theory Econ Manag*. (2016) 98–112. doi: 10.3969/j.issn.1000-596X.2016.02.008
56. Cleary S. The relationship between firm investment and financial status. *J Finance.* (1999) 54:673–92. doi: 10.1111/0022-1082.00121
57. Fama EF, French KR. Common risk factors in the returns on stocks and bonds. *J Finance Econ.* (1993) 33:3. doi: 10.1016/0304-405X(93)90023-5
58. Denis DJ, Osobov I. Why do firms pay dividends? International evidence on the determinants of dividend policy. *J Financ Econ.* (2008) 89:62–82. doi: 10.1016/j.jfineco.2007.06.006
59. Steven AB, Dong Y, Corsi T. Global sourcing and quality recalls: an empirical study of outsourcing-supplier concentration-product recalls linkages. *J Oper Manag.* (2014) 32:241–53. doi: 10.1016/j.jom.2014.04.003
60. Casalin F, Pang G, Maioli S, Cao T. Inventories and the concentration of suppliers and customers: evidence from the Chinese manufacturing sector. *Int J Prod Econ.* (2017) 193:148–59. doi: 10.1016/j.ijpe.2017.07.010
61. Wang Z, Wang Z. The working capital management based on vendor relationship: "Icing on the Cake" or "Timely Help". *Nankai Bus Rev*. (2017) 20:32–44. doi: 10.3969/j.issn.1008-3448.2017.02.004
62. Xu S, Xu L. Cash dividend policy, a-gency cost and corporate performance. *J Manag Sci*. (2015) 28:96–110. doi: 10.3969/j.issn.1672-0334.2015.01.008
63. Meng Q, Bai J. Customer concentration and enterprise technological innovation: help or hinder – research based on individual characteristics of customers. *Nankai Manag Rev*. (2018) 4:62–73.
64. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* (1986) 51:1173. doi: 10.1037/0022-3514.51.6.1173
65. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociol Methodol.* (1982) 13:290–312. doi: 10.2307/270723
66. Wen Z, Ye B. Analyses of mediating effects: the development of methods and models. *Adv Psychol Sci.* (2014) 22:731. doi: 10.3724/SP.J.1042.2014.00731
67. Li S. Property rights, agency costs and agency efficiency. *Econ Res*. (2007) 42:102–13.
68. Jiang F, Huang L, ZHang M. Product market competition, corporate governance and agency costs. *World Econ.* (2009) 32:46–59.
69. Hadlock CJ, Pierce JR. New evidence on measuring financial constraints: moving beyond the Kz index. *Rev Financ Stud.* (2010) 23:1909–40. doi: 10.1093/rfs/hhq009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/fpubh.2022.909241

A Study on the Effect of the Pre-Go-Live Training in Anxiety and Depression of Medical Staff Based on the Data of Wuhan Fangcang Shelter Hospital During COVID-19 in the Era of Big Data

Qiang Feng[1,2], *Huizhi Zhou*[1], *Lu Wang*[1] *and Chuanyuan Kang*[1]*

[1] *Department of Psychosomatic Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China,*
[2] *Tongji University School of Medicine, Shanghai, China*

Coronavirus disease 2019 (COVID-19) broke out in 2019. In the past 4 years, China has adopted many measures to control the epidemic, including building Fangcang shelter hospitals to isolate confirmed positive cases. Therefore, we aim to explore the mental health status of medical staff in the Wuhan Fangcang shelter hospital and discuss the relevant factors that affect the medical staff's mental status. The subjects of the research were staff from several Fangcang shelter hospitals in Wuhan during the epidemic of COVID-19. Patient Health Questionnaire−9 items Scale (PHQ-9) was used to assess the severity of the participants' depressive symptoms, and Generalized Anxiety Disorder−7 items Scale (GAD-7) was used to evaluate the severity of the participants' anxiety symptoms. The demographic information and health adjustment methods were collected in a self-made questionnaire, and regression analysis on related factors that affect mental health was performed. The three most frequently used methods of psychological adjustment for the staff in the Fangcang shelter hospital are common recreational activities, such as reading, streaming videos, listening to music, and playing games. (93.8%), communicating with colleagues in the Fangcang shelter hospital (92.5%), and communicating with family members and friends (78.3%). Binary logistic regression analysis showed that developing depression symptoms has relation to 2 factors, which are having not participated in medical emergency rescue missions (odds ratio = 2.610; 95% confidence interval 1.398–4.872, $P = 0.003$) and inadequate training before entering the shelter hospital (odds ratio = 2.804, 95% confidence interval 1.293–6.08, $P = 0.009$). Compared with adequate pre-job training, insufficient training increases the risk of anxiety symptoms (odds ratio = 2.692; 95% confidence interval 1.3–5.575, $P = 0.008$). Lack of experience and inadequate training in medical emergency rescue missions exposed the medical staff to a higher risk of developing symptoms of depression and anxiety. Psychological adjustment methods that are helpful to adjust their mental state are most commonly used.

Keywords: COVID-19, Fangcang shelter hospital, mental health care, health psychology, public health

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Navod Thilakarathne,
University of Colombo, Sri Lanka
B. V. Santhosh Krishna,
New Horizon College of Engineering (NHCE), India

***Correspondence:**
Chuanyuan Kang
kangbao98@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *31 March 2022*
Accepted: *12 April 2022*
Published: *31 May 2022*

Citation:
Feng Q, Zhou H, Wang L and Kang C (2022) A Study on the Effect of the Pre-Go-Live Training in Anxiety and Depression of Medical Staff Based on the Data of Wuhan Fangcang Shelter Hospital During COVID-19 in the Era of Big Data. Front. Public Health 10:909241. doi: 10.3389/fpubh.2022.909241

INTRODUCTION

With the development of technology, big data has been more frequently used in people's life. By the end of December 2019, the coronavirus pneumonia (COVID-19) (1) was spreading domestically in Wuhan city of China, which quickly became the epicenter of the COVID-19 pandemic. Under the circumstance of shortage of hospital beds, Wuhan Fangcang shelter hospitals were designed and provided for the first time in China to tackle the COVID-19 outbreak and offer medical care for patients with mild or moderate COVID-19 (2). Within a few months, Wuhan opened a whopping 16 Fangcang shelter hospitals equipped with over 13,000 hospitals beds (3).

Under such a crisis situation, healthcare providers on the frontline caring for patients positive for COVID-19 were prone to developing psychological distress and mental symptoms. Recent research shows that medical workers caring for patients with COVID-19, especially nurses and women, in fever clinics and wards in Wuhan and other regions in China reported experiencing more psychological stress (4), and that they experienced the challenges of daily clinical work in COVID-19 wards (5–7). As the epidemic of COVID-19 progressed, local healthcare workers in Wuhan became overworked under the situation of the accumulative number of suspected and confirmed clinical cases in the process of the outbreak. In response to the medical crisis, 42,000 healthcare workers (including 28,600 nurses) across pan-China came to Wuhan's aid, including neighboring cities within Hubei province (8). Although it was a joint effort in an unfamiliar environment full of challenges, medical team members from their respective hospitals are kept intact during the Fangcang shelter tenure (9, 10). In addition, professional support from psychiatrists who consistently conducted face-to-face group psychological interventions and online psychological services for medical staff have a positive impact on maintaining mental health (11).

Therefore, we explored the mental health outcomes of healthcare workers caring for patients with COVID-19 in Fangcang shelter hospitals in Wuhan. To improve the validity of our research, we made a questionnaire to investigate 240 patients from the Wuhan hospital. We also provided additional evidence with respect to the mental wellbeing of healthcare workers. In addition, we used Patient Health Questionnaire-9 items Scale (PHQ-9) to evaluate the severity of the participants' depressive symptoms and Generalized Anxiety Disorder-7 items Scale (GAD-7) to assess the severity of the participants' anxiety symptoms.

METHOD

Participants and Sampling

This study was cross-sectional designed and conducted between February and March 2020, ∼1.5 months after the novel coronavirus disease 2019 (COVID-19) outbreak. A total of 240 participants recruited were medical workers from Fangcang shelter hospitals located in Wuhan city in China. The participants were allowed to finish the survey after the closure of Fangcang shelter hospitals. A total of 240 medical staff members completed the questionnaires voluntarily. The average working time of the participants in the Fangcang shelter hospital was 21.73 ± 10.69 days, 189 people (78.8%) were from Wuhan Living Room Fangcang Shelter Hospital, and 51 (21.2%) were from other Fangcang shelter hospitals. The average age of the participants was 35.27 ± 7.31 years old, including 159 women (66.3%), 81 men (33.8%), 82 doctors (34.2%), 143 nurses (59.6%), and 15 other staff (6.3%). Of 189 people, 58 individuals (24.2%) have previously participated in medical emergency rescue missions. In terms of mental health, 83 people (34.6%) had depression symptoms, and 93 people (38.8%) had anxiety symptoms. Univariate analysis showed that people who had not participated in medical emergency rescue missions before were more prone to depression symptoms ($\chi^2 = 8.035$, $P < 0.05$). The healthcare workers with insufficient training before entering Fangcang shelter hospitals had a greater proportion of depression and anxiety symptoms ($P < 0.05$).

In accordance with the Declaration of Helsinki, the ethics panel of the Medical Association of Shanghai East Hospital approved this study (No. 2020007). All the participants provided informed consent for research details and anonymized responses to be published. They were clearly informed that they had the right to refuse or withdraw from the research at any time. In the light of standardized instructions, we introduced the research purpose, questionnaire content, study process, and principle of confidentiality to the potential participants. The survey was anonymous, and because of the requirements of epidemic prevention, informed consent was provided electronically by all the survey participants after participating in the study.

Assessment Procedure

The survey participants used a mobile phone or a tablet to scan a WeChat QR code to access the mobile questionnaire survey system. All the participants filled in the questionnaires and scales online, which consisted of a general information questionnaire, a questionnaire on Fangcang shelter hospital work, the 9-item Patient Health Questionnaire Scale (PHQ-9), and the 7-item Generalized Anxiety Disorder Scale (GAD-7).

Questionnaire

This questionnaire included age, gender, occupational role (physician, nurse, or other), marital status, educational level (<undergraduate, bachelor or >graduate), types of personality traits, and past history of mental illness. This questionnaire was self-reported by the participants, including items related to the Fangcang shelter hospitals' scope of work, such as training and mental preparation prior to entering hospital, stress factors before work, past experience in emergency rescue, and coping methods.

Depression and Anxiety Assessment

With Chinese versions of validated measurement tools, symptoms of depression and anxiety among all the participants are investigated. As such, PHQ-9 and GAD-7 were used to assess the degree of severity of depression and anxiety symptoms, respectively. The outcome scores of measurement tools were interpreted as follows: the two scales' categories were both based

on established values in the literature, and items were rated on a 4-point Likert-type scale, ranging from 0 (not at all) to 3 (nearly every day). Total scores of PHQ-9 range from 0 to 27, with high scores meaning more depression symptoms. Based on existing original validation studies, the outcome of total scores was interpreted as four symptom levels suggesting no depression (0–4), mild (5–9), moderate (10–14), and severe (15–27), separately. Total scores of GAD-7 range from 0 to 21, with high scores meaning more anxiety symptoms of four levels of classification, which suggest no anxiety (0–4), mild (5–9), moderate (10–14), and severe (15–21), separately.

Investigation on Psychological Adjustment Methods

Nine self-help psychological adjustment methods were investigated for the participants, including communication with family members and friends, communication with colleagues in the Fangcang shelter hospital, physical exercise, typical recreational activities (such as reading, watching videos, listening to music, and playing games), reading books that guide psychological adjustment (either paper or Ebook), listening to an audio guide to psychological adjustment, self-learning self-relaxation techniques (such as breathing adjustments and progressive muscle relaxation exercises) through the Internet, calling the psychological assistance hotline, and seeking help from a psychological specialist (online or offline).

Statistical Analysis

SPSS 22.0 (IBM, Chicago) was used for statistical analysis. The significance level was set at $\alpha = 0.05$, and all tests were 2-tailed. Descriptive statistics were computed for categorical variables, and a Chi-squared test compared their respective rate variation. The continuous data of this questionnaire were not normally distributed and therefore presented as medians with interquartile ranges (IQRs). The non-parametric Mann-Whitney U test and the Kruskal-Wallis test were conducted to compare the severity of symptoms between two or more groups. To explore potential risk factors for symptoms of depression and anxiety in the participants, a logistic regression analysis was performed, and the risk factors are presented as odds ratios (ORs) and 95% confidence intervals (CIs), with adjustment for confounders, namely, age, gender, marital status, educational level, emergency rescue experience, and degree of training.

RESULT

In the study, 240 participants volunteered to complete the survey, 82 were physicians, 143 were nurses, and 15 were miscellaneous staff (such as laboratory personnel). Majority of the participants were women (159) aged 30–40 years (134), were married (180), had an educational level of undergraduate (153), and worked for 2–4 weeks in the Fangcang shelter hospital (**Table 1**).

Physicians vs. Nurses and Subgroup Analysis

Differences between the physicians and the nurses in terms of demographic characteristics, namely, gender, age, education level, marriage, and length of work, were statistically significant ($P < 0.05$). Most of the doctors are male, older, married, and have higher education levels and longer working years.

According to this study, the physicians and nurses working in Fangcang shelter hospitals vary in terms of personal background. The physicians (31.7 vs. 17.5%) had more experience in rescue missions than the nurses ($P < 0.05$). However, the nurses trumped the physicians in terms of support from family members (66.4 vs. 48.8%) to work in Fangcang shelter hospitals. Among the physicians, highly educated physicians were more likely to worry about being infected (85.4 vs. 61%); young physicians are more worried about being incapable of working in the Fangcang shelter hospital and having a heavy workload ($P < 0.05$).

Among physicians, those who had an above bachelor's degree were more likely to be stressed for the probability to be infected. Speaking to family was the method of choice of more educated doctors for psychological self-adjustment compared with the counterpart with less academic education. Physicians with above bachelor's degrees showed higher rates of depression and anxiety symptoms (**Table 2**). There were no similar results among the nurses.

With Past Emergency Rescue Experience vs. Without Past Emergency Rescue Experience

Compared with those who had emergency rescue experience, participants with no emergency rescue experience were more nervous about overload (before they went to Fangcang shelter hospital [38 (20.9%) vs. 5 (8.6%), $p = 0.03$]. Those with emergency rescue experience were more worried about being infected ($Z = -3.557$, $p < 0.05$ and the health of their families ($Z = -3.056$, $p < 0.05$) during their Fangcang shelter hospital tenure. The difference between PHQ-9 and GAD-7 scores was not statistically significant.

Risk Factors of Depression and Anxiety

The logistic regression analysis identified the following two groups to be at a higher risk of developing depression symptoms: no emergency rescue experience and inadequate training. In addition, the logistic regression analysis showed that the individuals with inadequate training were at higher risk of experiencing anxiety symptoms (**Table 3**).

Psychological Adjustment Methods and Self-Evaluation Effectiveness

The top three frequently used methods of psychological adjustment for the staff in the Fangcang shelter hospitals are usual entertainment (such as reading, watching videos, listening to music, and playing games) (93.8%), communicating with colleagues in the Fangcang shelter hospital (92.5%), and communicating with family members and friends (78.3%), and the three most effective methods of psychological adjustment were calling the psychological assistance hotline (81.8%), seeking help from a psychological specialist (64.3%), communicating with colleagues in the Fangcang shelter hospital 120 (54.1%) (**Table 4**).

TABLE 1 | Demographic characteristics of participants.

Variable N	Total (%)	Physicians (%)	Nurses (%)	Other staffs (%)
Overall	240 (100)	82 (34.2)	143 (59.6)	15 (6.3)
Gender				
Men	81 (33.8)	57 (69.5)	13 (9.1)	11 (73.3)
Women	159 (66.3)	25 (30.5)	130 (90.9)	4 (26.7)
Marriage status				
Unmarried	60 (25.0)	9 (11.0)	48 (33.6)	3 (20)
married	180 (75)	73 (89.0)	95 (66.4)	12 (80)
Education level				
<Undergraduate	40 (16.7)	2 (2.4)	36 (25.2)	2 (13.3)
=bachelor	153 (63.8)	39 (47.6)	104 (72.7)	10 (66.7)
>Graduate	47 (32.9)	41 (50.0)	3 (2.1)	3 (20.0)
Age (years)				
<30	55 (22.9)	5 (6.1)	46 (32.2)	4 (26.7)
30–40	134 (55.8)	43 (52.4)	85 (59.4)	6 (40.0)
>40	51 (21.3)	34 (41.5)	12 (8.4)	5 (33.3)
Length of work (years)				
≤10	115 (47.9)	31 (37.8)	79 (55.2)	5 (33.3)
11–20	80 (33.3)	22 (26.8)	49 (34.3)	9 (60.0)
>20	45 (18.8)	29 (35.4)	15 (10.5)	1 (6.7)
Weeks in the Fangcang shelter hospital				
≤2	73 (30.4)	26 (31.7)	37 (25.9)	10 (66.7)
2–4	114 (47.5)	29 (35.4)	83 (58.0)	2 (13.3)
>4	53 (22.1)	27 (32.9)	23 (16.1)	3 (20)
Emergency rescue experience				
Yes	58 (24.2)	26 (31.7)	25 (17.5)	7 (46.7)
No	182 (75.8)	56 (68.3)	118 (82.5)	8 (53.3)

TABLE 2 | Subgroup physicians ($n = 82$).

Category *N* (%)	Undergraduate and below $n = 41$	Above bachelor's degree $n = 41$	*p*
Worries being infected before work	25 (61.0)	35 (85.4)	0.013
Talk to family for psychological adjustment	26 (63.4)	36 (87.8)	0.010
Depression	10 (24.4)	21 (51.2)	0.012
Anxiety	9 (22.0)	18 (43.9)	0.034

**Fisher exact method.*

TABLE 3 | Risk factors for depressive and anxiety symptom.

Variable	OR	95%CI	*p*
PHQ-9			
No emergency rescue experience	2.610	1.398–4.872	0.003
Inadequate training	2.804	1.293–6.080	0.009
GAD-7			
Inadequate training	2.692	1.300–5.575	0.008

**Binary logistic regression.*

DISCUSSION

This study found that there was no variance in anxiety and depression levels between the physicians and the nurses. Previous research shows that women, nurses, frontline healthcare workers, and those working in Wuhan reported significantly higher severity of mental health symptoms than all other healthcare workers. The frontline healthcare workers who engaged in the daily work of diagnosis, treatment, and nursing of patients with COVID-19 had a higher risk of experiencing symptoms of insomnia, anxiety, depression, and distress (4). Other research shows that being at risk of contact with patients with COVID-19, living in rural areas, and being female were the most common risk factors for people developing insomnia, anxiety, depression, and obsessive-compulsive symptoms (5). The possible rationale might be fear of infection, prolonged working hours, lack of protective medical apparatus, infected patient load, less effective COVID-19 medication, loss and death of medical colleagues after exposure to COVID-19, social distancing, or isolation from family and friends, and the deteriorating situation of patients may have a negative impact on the mental health of medical workers (12).

TABLE 4 | Psychological adjustment method and Self-evaluation effectiveness.

Psychological adjustment method	Number of options *n* (%)	Self-evaluation effectiveness *n* (%)			
		Very helpful	Helpful	Help a bit	Helpless
1. Communicating with family members and friends	188 (78.3)	99 (52.7)	71 (37.8)	15 (8.0)	3 (1.6)
2. Communicate with colleagues in the Fangcang shelter hospital	222(92.5)	120 (54.1)	82 (36.9)	14 (6.3)	6 (2.7)
3. Physical exercise	168 (70.0)	78 (46.4)	71 (42.3)	12 (7.1)	7 (4.2)
4. Usual entertainment	225 (93.8)	105 (46.7)	103 (45.8)	12 (5.3)	5 (2.2)
5. Read books that guide psychological adjustment	132 (55.0)	63 (47.7)	58 (43.9)	8 (6.1)	3 (2.3)
6. Listen to an audio guide to psychological adjustment	92 (38.3)	46 (50.0)	36 (39.1)	9 (9.8)	1 (1.1)
7. Self-learning self -relaxation techniques	110 (45.8)	53 (48.2)	45 (40.9)	9 (8.2)	3 (2.7)
8. Call the psychological assistance hotline	11 (4.6)	9 (81.8)	0	1 (9.1)	1 (9.1)
9. Seek help from a psychological specialist	14 (5.8)	9 (64.3)	4 (28.6)	0	1 (7.1)

Compared with those of the physicians, the nurses' families were more supportive of their work in the Fangcang shelter hospital [40 (48.8%) vs. 95 (66.4%), $p = 0.03$]. There were no differences in scores measuring symptoms of depression and anxiety between the two groups. Those who had experiences of previous rescue missions were more worried about getting infected and the health status of their family members. By regression analysis, it was found that depression and anxiety symptoms were related to whether they had participated in rescue training and sufficient pre-go-live training.

Research shows receiving negative information about COVID-19 and participating in frontline work against COVID-19 appear to be vital risk factors for mental health problems (13–15). In terms of psychological adjustment, this study demonstrated that doctors with higher education levels had more contacts with their family members (87.8 vs. 63.4%); older doctors were more inclined to adjust their psychological pressure by reading ($Z = -1.99$, $P < 0.05$). Among the nurses, even the older nurses were more worried about being infected ($Z = -2.076$, $P < 0.05$), but they were more inclined to adjust their psychological stress through physical exercise ($Z = -3.001$, $P < 0.05$). Based on these investigation results, healthcare workers have sufficient resources at their disposal to overcome the stress related to their daily work in the Fangcang shelter hospital (16).

This result shows that within a centralized isolation treatment environment of the Fangcang shelter hospital, stress is widely existent, and depression and anxiety symptoms are common. For the medical staff in Fangcang shelter hospitals, training before on-boarding, appropriate recreational activities, and strengthening interpersonal communication and support can help reduce the stress, anxiety, and depression levels.

CONCLUSION

In conclusion, we found that, for the mental health of medical personnel fighting COVID-19, it is very important for them to have medical rescue experience and undergo pre-go-live training. With the psychological adjustment methods of improving interpersonal contact and professional help and peer support, medical workers can regain their psychological strength against stress from COVID-19.

However, this research is limited in time frame and because of the cross-sectional design. The psychological assessment was based on a voluntary online survey and self-report tools and had a small sample size. Future longitudinal studies are required for follow-up of the mental wellbeing of medical workers. Qualitative research interviews are also encouraged in future studies to generate a more comprehensive understanding of follow-up mental wellbeing adjustment (e.g., Balint group).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The ethics panel of the Medical Association of Shanghai East Hospital approved this study (No.2020007). All the participants provided informed consent for research details and anonymized responses to be published. They were clearly informed that they had the right to refuse or withdraw from the research at anytime. In the light of standardized instructions, we introduced the research purpose, questionnaire content, study process, and principle of confidentiality to the potential

participants. The survey was anonymous, and because of the requirements of epidemic prevention, informed consent was provided electronically by all the survey participants after participating in the study.

AUTHOR CONTRIBUTIONS

QF and CK: conception and design. QF, HZ, and LW: data analysis and interpretation. All authors: administrative support, provision of study materials or patients, collection and assembly of data, manuscript writing, and final approval of manuscript.

FUNDING

This study was supported by the Special Fund for Fighting the COVID-19 Outbreak sponsored by Tongji University School of Art and Media and Institute of Disaster Medicine, East Hospital, Tongji University (20TJBXKY06).

REFERENCES

1. China NHCo. *New Coronavirus Pneumonia Diagnosis and Treatment Plan (Trial version 7)*. (2020). Available online at: http://wwwnhcgovcn/yzygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989/files/ce3e6945832a438eaae415350a8ce964pdf (accessed June 17, 2021).
2. Chen S, Yang J, Yang W, Wang C, Bärnighausen T. COVID-19 control in China during mass population movements at New Year. *Lancet*. (2020) 395:764–66. doi: 10.1016/S0140-6736(20)30421-9
3. The State Council of the People's Republic of China. *The Joint Prevention and Control Mechanism of the State Council Launch Announcement on Further Shouldering Responsibilities and Implementing Prevention and Control Strategies (in Chinese)*. (2020). Available online at: http://www.gov.cn/guowuyuan/2020-02/07/ content_5475951.htm (accessed February 24, 2020).
4. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Network Open*. (2020) 3:e203976. doi: 10.1001/jamanetworkopen.2020.3976
5. Zhang WR, Wang K, Yin L, Zhao WF, Xue Q, Peng M, et al. Mental health and psychosocial problems of medical health workers during the COVID-19 epidemic in China. *Psychother Psychosom*. (2020) 89:242–50. doi: 10.1159/000507639
6. WHO. *WHO Coronavirus Disease (COVID-19) Situation Dashboard*. (2021). Available online at: https://covid19whoint/ (accessed June 17, 2021).
7. WHO. *Coronavirus Disease (COVID-19)*. (2021). Available online at: https://wwwwhoint/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses (accessed June 17, 2021).
8. Liu Q, Luo D, Haase JE, Guo Q, Wang XQ, Liu S, et al. The experiences of health-care providers during the COVID-19 crisis in China: a qualitative study. *Lancet Global Health*. (2020) 8:E790–8. doi: 10.1016/S2214-109X(20)30204-7
9. Chinanews. *Beijing: Two Departments Issue the Novel Coronavirus Infection Pneumonia Diagnosis and Treatment Plan. (trial version 7)*. Available online at: http://www.chinanews.com/gn/2020/03-04/9113100.shtml. Chinese (accessed March 4, 2020).
10. SINAnews. *Beijing: In January, Hubei Had More Than 3,000 Medical Infections, and the Wuhan Health and Medical Committee Reported "None" for Half a Month*. Available online at: https://news.sina.com.cn/o/2020-03-06/docIimxyqvz8395569.shtml. Chinese (accessed March 6, 2020).
11. Wang Y, Zhao X, Feng Q, Liu L, Yao Y, Shi J. Psychological assistance during the coronavirus disease 2019 outbreak in China. *J Health Psychol*. (2020) 25:135910532091917. doi: 10.1177/1359105320919177
12. *Mental Health and Coping During COVID-19*. Centers for Disease Control and Prevention (2020). Available online at: https://www.cdc.gov/coronavirus/2019-ncov/about/coping.html (accessed May 03, 2020).
13. Que J, Le Shi JD, Liu J, Zhang L, Wu S, Gong Y, et al. Psychological impact of the COVID-19 pandemic on healthcare workers: a cross-sectional study in China. *Gen Psychiatry*. (2020) 33:e100259. doi: 10.1136/gpsych-2020-100259
14. Fang D, Pan S, Li Z, Yuan T, Jiang B, Gan D, et al. Large-scale public venues as medical emergency sites in disasters: lessons from COVID-19 and the use of Fangcang shelter hospitals in Wuhan, China. *BMJ Global Health*. (2020) 5:e002815. doi: 10.1136/bmjgh-2020-002815
15. Dai LL, Wang X, Jiang TC, Li PF, Wang Y, Wu SJ, et al. Anxiety and depressive symptoms among COVID-19 patients in Jianghan Fangcang Shelter Hospital in Wuhan, China. *PLoS ONE*. (2020) 15:e0238416. doi: 10.1371/journal.pone.0238416
16. National Health Commission of the People's Republic of China. *Press conference of the joint prevention and control mechanism of the State Council on Feb 29, 2020*. National Health Commission of the People's Republic of China (2020). Available online at: http://www.nhc.gov.cn/xwzb/webcontroller.do?titleSeq=11248&gecstype=1 (accessed March 1, 2020).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Feng, Zhou, Wang and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fpubh.2022.922574

The Effect of Big Data-Based Digital Payments on Household Healthcare Expenditure

Chengming Li[1], *Daming Li*[1], *Si He*[1], *Shiqi Sun*[2], *Yuan Tian*[3] *and Zeyu Wang*[4]*

[1] School of Economics, Minzu University of China, Beijing, China, [2] School of Software & Microelectronics, Peking University, Beijing, China, [3] Center for Enterprise Growth and National Economic Security Research, Tsinghua University, Beijing, China, [4] School of Public Administration, Guangzhou University, Guangzhou, China

Whether families using big data-based digital payments will increase household healthcare expenditure is a subject that needs to be investigated in the era of big data. Based on the data from China Family Panel Studies (CFPS), 24,126 samples from 2014 to 2018 are used to examine the impact and mechanism of big data-based digital payments on household healthcare expenditure. The empirical results of this paper show that the use of digital payments by households can significantly increase household healthcare expenditure with the empowerment of big data. This research employs the instrumental variable method to verify and produce consistent estimation results in order to address potential endogeneity issues such as measurement error and missing variables. We learn via mechanism analysis that household adoption of big data-driven digital payments can remove credit limitations and build social capital, resulting in higher household health-care spending. We also perform a heterogeneity analysis. The findings reveal that when a family's traditional financial accessibility is high, the head of the household is young or middle-aged, and the head of the household has a higher level of education, digital payment will play a larger role in encouraging household healthcare expenditure. The conclusions of this paper are still solid after changing the indicators of household healthcare expenditure substituting the indicators of digital payment, and adjusting the variables. As a result, this article provides micro-evidence for the usage of digital payments by households to enhance healthcare spending.

JEL Classification: D12 G21 O30 O53 I12

Keywords: big data, digital payments, household healthcare expenditure, credit constraints, social capital

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Sri Vidhya,
Sathyabama Institute of Science and Technology, India
Roshan Fernandes,
Nitte Mahalinga Adyanthaya Memorial Institute of Technology, India

***Correspondence:**
Zeyu Wang
theowang@whu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *18 April 2022*
Accepted: *02 May 2022*
Published: *02 June 2022*

Citation:
Li C, Li D, He S, Sun S, Tian Y and Wang Z (2022) The Effect of Big Data-Based Digital Payments on Household Healthcare Expenditure. Front. Public Health 10:922574. doi: 10.3389/fpubh.2022.922574

INTRODUCTION

Household healthcare expenditure have risen to become the third greatest family expense after housing and education, and the phenomenon has been attracting the attention of many academics. Economic strife (1), family income (2–6), age and gender (7, 8), microfinance (9, 10), and other factors have been identified to influence household healthcare expenditure in previous studies.

Health problems are growing more serious as China's economy grows, and households are paying greater attention to medical treatment. The share of families working overtime and staying up late in Beijing is substantially higher than the general average level, according to the survey results of 2021 Chinese Urban Family Health Life Power White Paper. Families in

Beijing and Shanghai are burdened by a greater economic load, while families in Zhengzhou are plagued by mental illness and financial strain. Families are increasingly confronted with acute mental sub-health and intense pressure in an era when individuals are more concerned about their health. Health security is becoming an increasingly significant aspect of family health management. For example, 65 percent of families agree that customizing commercial health insurance is vital. When compared to traditional payment methods, digital payment based on big data is more favorable to breaking down time and space barriers and lowering the threshold for consumers to enjoy financial services, allowing them to do so more conveniently and swiftly. Digital medicine has become an active choice to alleviate the shortage of medical resources and limited medical conditions, thanks to the empowerment of digital technologies like big data, cloud computing, and blockchain. Digital payment has a bigger role to play in the growth of digital healthcare. As a result, from the standpoint of big data, this research investigates the effect of big data-based digital payments on household healthcare expenditure.

On the one hand, households' use of digital payment based on big data can broaden available financial services, increase sources of income, and alleviate credit constraints; on the other hand, we can accumulate social capital, broaden information channels, and reduce transaction costs, which may lead to an increase in household healthcare expenditure. However, existing research does not provide a clear response to the question of how digital payments based on big data affect household health care spending. The majority of studies look at the impact of family income (6), age and gender (8), and microfinance (9) on family healthcare expenditure, or address the potential benefits and risks of digital payment (11) and the impact of digital payment on rural household consumption in China (12, 13). Of course, some articles focus specifically on the impact of mobile money on household medical spending (14). However, there are few papers that look at the influence of digital payments on household health care spending from a big data viewpoint.

As a result, utilizing the China Family Panel Studies (CFPS) database and 24,126 family sample data from 2014 to 2018, this research investigates the influence of digital payment on household healthcare expenditure from a big data viewpoint, yielding the following findings. First, this research shows that digital payment enabled by big data can boost household health care expenditures using OLS estimation. Second, we use the instrumental variable technique to verify and get consistent estimation results in order to address potential endogenous issues such as measurement errors and missing data. Third, the mechanism analysis reveals that digital payments enabled by household big data applications can reduce credit limitations and build social capital, resulting in an increase in household health-care spending. Fourth, the heterogeneity analysis reveals that the use of digital payment based on big data is more effective in encouraging household healthcare expenditure in families headed by middle-aged and young individuals with a high level of education and financial accessibility. The conclusions of this research are still solid after replacing the indicators of household healthcare expenditure, replacing the indicators of digital payment, and tail reduction. As a result, this article presents micro evidence encouraging families to adopt digital payment to boost their household healthcare expenditure.

The following are the contributions of this paper. First, this paper adds to the body of knowledge about the factors that influence household healthcare expenditure (5, 6, 8, 9). The existing literature focuses mostly on home consumption studies and does not consider the influence of household health care spending. Song et al. (15) and Li et al. (16), for example, looked at the relationship between digital finance and household consumption from a digital finance perspective; Hou et al. (17) believed that liquidity constraints would affect household consumption; Kang (18) believed that social networks would increase household consumption. Furthermore, a few studies have looked into the characteristics that influence household healthcare expenditure. Microfinance (9), family income (5, 6), and age and gender (8) all have an impact on household healthcare expenditure. This study investigates the influence of digital payment based on big data on household healthcare expenditure and clarifies the impact of payment mode shift on household healthcare expenditure in the digital era through empirical analysis.

Second, this research adds to the body of knowledge about the influence of digital payments on household spending. The majority of existing work focuses on the possible benefits and risks of digital payment (11), the influence of digital payment on Chinese households' overall spending (12, 13), etc. There are, of course, certain studies that look specifically at the influence of mobile money on household health-care spending (14). However, there is limited literature on the influence of digital payment on household healthcare expenditure under the empowerment of big data from a big data perspective. As a result, this work adds to the body of knowledge about the influence of digital payments on household consumption, as well as offering new perspectives and research ideas in this area.

Third, this study adds to the body of knowledge about the impact of Fintech on family economic behavior. The impact of consumer finance (19, 20) and technology finance (21–23) on family economic behavior is the subject of most existing studies. Digital payment is an important financial technology tool, but there is little research on its impact on family healthcare behavior. This research provides new evidence regarding the impact of Fintech on family economic behavior based on this, from the standpoint of digital payment, along with the enabling function of big data.

This paper's structure is as follows: The theoretical analysis of digital payment on household healthcare expenditure under big data empowerment is summarized in Section Theoretical Analysis and Research Hypotheses; Section Data and Methods provides the data, variables, and models used in empirical analysis; Section Results presents the major result, which includes a mechanism and robustness analysis; Section Further Discussion: The Heterogeneity of Household Healthcare Expenditures presents the paper's further analysis; and Section Conclusion presents the conclusion.

THEORETICAL ANALYSIS AND RESEARCH HYPOTHESES

Household Use Big Data-Based Digital Payments Can Ease Credit Constraints and Thus Increase Household Healthcare Expenditure

Financial institutions gain access to a large amount of household data after households use big data-based digital payments. On the one hand, financial institutions can use big data from households to understand households' consumption preferences and accurately deliver personalized information by mining households' consumption data in order to increase households' demand for financial services (24); on the other hand, financial institutions can use big data analytics (BDA) to build decentralized and distributed models to predict households' credit default probabilities (25). As a result, big data-driven digital payments can affect both the supply and demand sides of financial services to ease household credit constraints encouraging households to use healthcare services (26).

On the demand side of financial services, households' use of big data-based digital payments helps them increase their financial and entrepreneurial incomes (12), while higher incomes increase the likelihood of successful credit access and help alleviate credit constraints. On the one hand, financial institutions use big data to push appropriate financial products to households based on their characteristics and needs in order to increase household financial income. Households can better manage their finances with the help of big data-based digital payments, such as purchasing money funds and investing in equity funds, thereby increasing their financial income. On the other hand, big data based-digital payments can stimulate households' entrepreneurial enthusiasm and thus increase their entrepreneurial income by providing functions such as information transfer, credit, and social interaction (27, 28).

On the supply side of financial services, on the one hand, financial institutions analyze households' credit, consumption, social, and other behaviors based on big data generated by digital payments (29) to accurately grasp households' credit status (30, 31) and include creditworthy but without collateral households into the scope of credit allocation and expand the coverage of financial service supply. Financial institutions, on the other hand, can use big data to better control the risk of credit disbursement, lower the cost of credit risk, and exit more low-threshold credit services. As a result, households can use microfinance services provided by digital platforms like WeChat and Alipay to gain access to more flexible, convenient, and low-threshold credit, easing their credit constraints (32). Traditional payment tools can achieve microfinance mobilization in the short term, but some households will face problems such as financial exclusion due to high barriers erected by financial institutions and a lack of awareness among some rural households, which will not be able to alleviate credit constraints (33–35).

In conclusion, households' use of big data-based digital payments can boost household healthcare expenditure by increasing income sources from both the demand and supply sides. As a result, this paper proposes hypothesis below.

H1: Households use big data-based digital payments to alleviate credit constraints and thus boost household healthcare expenditure.

Household Use Digital Payments Based on Big Data Can Increase Social Capital and Thus Increase Household Healthcare Expenditure

Digital platforms in China, such as WeChat and Alipay, offer not only digital payment services but also virtual social features. When households use big data-based digital payments, these digital platforms tend to use big data, based on a MapReduce-based paradigm, to perform social data analysis through the most popular framework, Apache Hadoop, to drive the creation and strengthening of social networks among different households on digital platforms to increase the viscosity of households' use of digital platforms to increase the viscosity of households' use of digital platforms to increase the viscosity of households' use of digital platforms (36). The formation and strengthening of social networks increase not only the viscosity of families' use of digital platforms, but also their social capital. In other words, households' use of big data-based digital payments on increases the frequency of communication with friends, relatives, and other subjects, thereby expanding and strengthening household social networks and contributing to the accumulation of social capital (37). As a result, increasing household social capital facilitates the expansion of information channels and the reduction of transaction costs, which increases household healthcare expenditure.

Households' use of big data-based digital payments expands access to information. Households are connected to digital platforms and their social capital grows when they use big data-based digital payments. Following that, digital platforms like WeChat and Alipay can push more healthcare-related information to households, increasing their access to information. Households can purchase products such as health insurance and health consultation based on information pushed to the Internet by big data. Furthermore, digital payments make it easier for households to buy healthcare-related products and services by streamlining the transaction process by allowing payments and transfers to be made via the Internet (22, 38).

Furthermore, households can reduce transaction costs in healthcare visits by using big data-based digital payments. When households build social capital through big data-based digital payments, they become more connected to subjects like family and friends, other households, and hospitals, giving them more opportunities and resources while lowering transaction costs (19, 39). Hospitals can use big data from households to analyze the health needs of households in real time and provide remote registration and online medical services as needed, establishing triage mechanisms and personal medical databases for patients and adjusting resource allocation, making it more efficient and convenient for patients to see a doctor and, to some extent, reducing the transaction costs of household healthcare (40). It is also easy for banks to assess the creditworthiness of potential borrowers using big data from households, allowing

them to provide more accurate and cheaper credit to households (41), lowering the transaction costs of credit between banks and households.

In conclusion, households' use of big data-based digital payments can reduce widening information channels, lower transaction costs, increase social capital, and thus boost household healthcare expenditure. Thus, this paper proposes hypothesis below.

H2: households' use of big data-based digital payments can accumulate social capital, boosting household healthcare expenditure.

Different Types of Household Applications of Big Data-Based Digital Payments Have Heterogeneous Impact on Household Health Care Expenditure

Furthermore, the use of big data-based digital payments by households has a heterogeneous effect on household health care expenditures, according to our study. Financial literacy is influenced by socio-demographic characteristics (e.g., age, education) in households (42), and because big data-based digital payments are a key component of digital finance, financial literacy influences households' use of big data-based digital payments. In terms of age differences, young and middle-aged people typically have higher cognitive levels, and higher cognitive levels significantly increase wage income (43, 44). Increased income allows this group to use big data-based digital payments more frequently, which boosts household income (26). Due to physical aging, the cognitive level of household heads declines as they grow older. As a result, young and middle-aged heads of households typically have a higher level of cognition than older heads of households, and are able to use and enjoy the benefits of big data-based digital payments more quickly, resulting in higher household healthcare expenditure. When it comes to education, the more educated group typically has better learning abilities and a larger knowledge base (45). Big data-based digital payments are novel and distinct from traditional payments, and they operate at a certain threshold, requiring users' ability to learn. Without assistance, the less educated may not be able to use or comprehend the features of big data-based digital payments, and may be hesitant to use them. As a result, higher education lowers the barrier to household heads using big data-based digital payments (46).

The use of big data-based digital payments is more prevalent in promoting household health care expenditures among households living in areas with greater financial accessibility on a regional level. Financial exclusion has long been a problem in rural areas of the country with distinct geographic characteristics (47), and the presence of financial institution outlets can effectively mitigate financial exclusion (47, 48). In other words, if households live in areas with limited access to traditional financial services, they are more likely to be financially excluded due to their remote location. Households can enjoy traditional financial services more conveniently and have good habits of participating in financial services in places with convenient transportation and high accessibility to traditional financial services, making them more likely to accept big data-based digital payments and play the role of big data-based digital payments in promoting household healthcare expenditure.

H3: The use of digital payments contributes more to household healthcare expenditure if the household has higher traditional financial accessibility, the household head is young or middle-aged, and the household head is more educated.

DATA AND METHODS

Model Setting

Referring to Liu et al. (49) and Zhang et al. (13), this paper develops an empirical model to explore the impact of digital payment use on household healthcare expenditure. The specific model is set up as follows.

$$Expenditure_{ijt} = \alpha_0 + \beta_0 Digital\ payments_{ijt} + \gamma_0 Control_{ijt} + \varphi_i + \delta_t + \varepsilon_{ijt} \quad (1)$$

The explanatory variable $Expenditure_{ijt}$ is household healthcare expenditure from the t th year and j th province and i th household. The household healthcare expenditure in the past year is obtained by logarithmic treatment. The core explanatory variable $Digital\ payments_{ijt}$ is the big data-based digital payment, which is represented by the frequency of Internet business activity of the household head in the usual case. We focus mainly on the coefficient β_0 to see the impact of digital payments on household healthcare expenditure. A positive coefficient indicates that big data-based digital payments boost household healthcare expenditure. If the coefficient is negative, it indicates that digital payments reduce household healthcare expenditure. $Control_{ijt}$ is the household head-level and household-level control variables, φ_i denote household fixed effects, and δ_t denote time fixed effects. Since households may be correlated in the same province, we cluster the standard errors to the provincial level.

Indicator Metrics

Explanatory Variable: Household Healthcare Expenditure

The explanatory variable is household healthcare expenditure. Healthcare expenditures are divided into two categories in the CFPS data: household medical expenditure and household fitness expenditure. Specifically, the CFPS questionnaire addresses household healthcare costs by asking, "What medical expenditure were paid directly by your household in the last 12 months? Of these, fitness expenditure do not include expenses that have been repaid or are expected to be reimbursed, but they do include certain expenses borrowed from friends and relatives or previously paid." "How much did your family pay for healthcare in the past 12 months? Healthcare expenditures include, among other things, the cost of physical activity and the purchase of related equipment and healthcare items." It can be observed that the first question indicates the household's medical care spending, while the second question reflects the household's healthcare spending, which combined show the significance of healthcare to the household. To get the logarithmized result, we add up the medical and healthcare costs paid directly by

households in the previous 12 months to get a total, then add 1 to the total and take the natural logarithm of the value after adding 1 to get the result.

Core Explanatory Variables: Digital Payments

The frequency of our household head's Internet business activity is used as a proxy variable for the extent to which the household uses digital payments. The household head is the primary decision-maker in the household, so the behavior of the household head has a significant impact on household activities. Specifically, this variable is addressed in the CFPS questionnaire by the question, "In general, how often do you use the Internet network for business activities (e.g., using Internet banking, online shopping)?" To provide a more visual representation of the extent to which households use digital payments, we assign a value to the frequency of the household head's response. We use 6 if the home head utilizes the Internet virtually every day for work purposes. We use 5 if the household head uses it 3 to 4 times a week; 4 if the household head uses it 1–2 times a week; 3 if the household head uses it 2–3 times a month; 2 if the household head uses it once a month; 1 if the household head uses it once every few months; and 0 if the household head never uses digital payments.

Other Control Variables

In terms of the selection of control variables, this paper draws on existing literature (12, 13, 23, 50–52) to control for both household-level influences and relevant variables at the household head- level. Specifically, at the household level, we control for household size, child support ratio, elderly support ratio, and net household income; at the household head level, this paper controls for the age of the household head and its squared term, the gender of the household head, the marital status of the household head, and whether the household head has party membership.

Specifically, the larger the household, the higher the household healthcare expenditure (53); the larger the proportion of children and the proportion of elderly, the higher the household's dependency burden, which may also contribute to household healthcare expenditure (54); the higher the household's net income, the higher the household's healthcare (6); and the age, gender, marital status, and party membership of the household head may also have a significant impact on household healthcare expenditures (8, 55).

Data Sources and Descriptive Statistical Analysis

The data we utilize comes from the China Family Panel Studies (CFPS) database, which is updated every 2 years. Because the most recent released household data from this survey are available up to 2018, and digital payments have been widely utilized since 2013, we selected to use data from 2014 to 2018. Referring to existing research methods (56), we processed the data as follows: (1) We eliminated samples with key codes missing, such as household, person, urban-rural, and provincial codes. (2) Since we used panel data and the distribution of data within each household is not necessarily the same, these data may show linearly varying relationships over the survey period. As a result, we used linear interpolation to fill in certain blanks (e.g., net household income per capita and age). Variables that were missing for three consecutive periods and could not be filled in were also eliminated. In addition, samples with incomplete gender and household size are excluded. This is because the household head's gender does not change, and changing the family size is difficult in the short term, therefore the interpolation technique is not applicable. In addition, we logarithmized variables such household expenditures and household net income (adding 1 to the values of the variables and taking the natural logarithm of the new values obtained after the addition of 1 to obtain the logarithmized results). (3) Since the CFPS does not give explicit information on the household head, researchers usually treat the principal, decision-maker, property owner, and financial respondent as the household head. However, since 2014, as information such as principal is no longer disclosed in the CFPS, only the financial respondent and the property owner in the household can be the head of the household, and most studies treat the financial respondent as the head of the household. Therefore, this paper identifies household heads by financial respondents in 2014 and only retains information on household heads who are 16 years old or older. (4) We logarithmized variables with large values such as income and distance. Specifically, we add 1 to their values and then take the natural logarithm to obtain the new variables. (5) For continuity in the time dimension, we only keep households that were surveyed in all 3 years. In addition, the nature of a household may fundamentally change when a household is split into multiple households or when a household moves, so in this paper we only retain households that were not split and did not move during the sample period.

After the above processing, we obtained 24,126 samples between 2014 and 2018, and the descriptive statistical results of the relevant variables are shown in **Table 1**. Among the household head and household control variables, the mean value of age of household head is 52.98, the standard deviation is 13.15, the minimum value is 16, and the maximum value is 93; the mean value of household size is 3.786, the standard deviation is 1.876, the minimum value is 1, and the maximum value is 21; the mean value of proportion of children is 0.124, the standard deviation is 0.201, the minimum value is 0, and the maximum value is 4; the mean value of elderly dependency ratio is 0.24. The above data indicate that the age distribution of household heads is wide, but the household structure is relatively similar.

RESULTS

Baseline Regression

We present the results of estimating the impact of digital payments on household healthcare expenditure in **Table 2**. Specifically, column (1) shows the regression results without any control variables, column (2) shows the regression results after controlling for household and time fixed effects, column (3) shows the regression results after adding household head-related variables to column (2), and column (4) shows the regression results after further adding household-related variables. As seen

TABLE 1 | Statistical description of variables.

Variable	Explanation	*N*	Mean	Sd	Min	Max
Healthcare expenditure	The sum of household medical expenditure and household health expenditure	24,126	6.901	2.711	0	13.84
Digital payments	Extent of digital payments used by the household head	24,126	0.460	1.276	0	6
Age of the household head	Age of the household head	24,126	52.98	13.15	16	93
Age^2 of the household head	$Age^2/100$ of the household head	24,126	29.80	14.13	2.560	86.49
Gender of the household head (male = 1)	Gender of the household head	24,126	0.514	0.500	0	1
Marriage of the household head (married = 1)	Marital status of the household head	24,126	0.870	0.336	0	1
Party membership (party = 1)	Whether the household head is a party member	24,126	0.107	0.309	0	1
Household size	Number of families eating together	24,126	3.786	1.876	1	21
Child ratio	Number of child in the household size	24,126	0.124	0.201	0	4
Elderly ratio	Number of elderly in the household size	24,126	0.240	0.350	0	3
Net household income	Net income is the income after deducting the cost of agricultural production	24,126	9.007	1.576	0	13.83

We define a child as an individual who is not older than 16 years old, an elderly person as an individual who is 60 years old and older.

TABLE 2 | Baseline regression results.

	(1)	(2)	(3)	(4)
Digital payments	0.0844***	0.0591**	0.0594**	0.0534**
	(5.90)	(2.36)	(2.19)	(2.07)
Age of the household head			−0.0170	−0.0410
			(−0.28)	(−0.61)
Square of age of the household head/100			0.0175	0.0387
			(0.36)	(0.79)
Gender of the household head			0.981	0.855
			(1.54)	(1.51)
Marriage of the household head			0.685***	0.554***
			(3.78)	(3.01)
Party membership			−0.0226	−0.0416
			(−0.11)	(−0.22)
Household size				0.224***
				(9.98)
Child ratio				−0.317
				(−1.55)
Elderly ratio				−0.0492
				(−0.34)
Net household income				0.0642***
				(3.32)
Household fixed effects	No	Yes	Yes	Yes
Time fixed effects	No	Yes	Yes	Yes
N	24,126	24,126	24,126	24,126
R^2	0.001	0.004	0.005	0.016

**** and ** indicate significant at the 1 and 5% statistical levels, respectively, with t-values in parentheses.*

in column (1), the regression coefficient of digital payment is 0.0844, which is significant at the 1% statistical level. As seen in columns (2), (3), and (4), this paper still shows a significant positive effect of digital payments based on big data on household healthcare expenditure after controlling for household and time fixed effects as well as the household head and household level influences.

Next, we observed the coefficients of the control variables and obtained the following results. The positive coefficient on marriage indicates that married heads of households value family protection and are more willing to spend on family healthcare; the coefficient on the number of household members is also positive, indicating that the larger the family size, the higher the household healthcare expenditure; and the coefficient on net household income is also significantly positive, indicating that households with higher income also value household healthcare, and increase expenditure more.

Robustness Analysis

Endogenous Processing

Due to omitted variables, this study may have endogeneity issues. In other words, variables other than those already controlled for in this paper may be associated with the use of digital payments or with households' health care expenditure, but we do not control for these other variables. In addition to this, there may be some error between the CFPS survey data and the actual household sentiment data, so there is also the possibility of measurement error in this paper.

Therefore, this paper uses the instrumental variable method to avoid the endogeneity problem. In this paper, the spherical distance from the capital city of the household's province to Hangzhou is taken as an instrumental variable. If Alipay originates in Hangzhou, the further the household is from Hangzhou, the less the use of digital payment. Therefore, we consider that the distance from the provincial capital city to Hangzhou is highly correlated with digital payment usage of households. In addition, the spherical distance is strongly exogenous and is not affected by digital payment use of households. The specific results are shown in **Table 3**.

The results show that the F-value of the weak instrumental variable test is >10 and passes the Wald test, indicating that

TABLE 3 | Robustness test: endogeneity treatment.

	(1)	(2)	(3)	(4)
	Digital payments	**Household healthcare expenditure**	**Digital payments**	**Household healthcare expenditure**
Digital payments		0.952***		1.046***
		(3.26)		(3.42)
Driving distance	−142.296 ***			
	(−3.66)			
Spherical distance			−1.145***	
			(−4.00)	
Control variables for household head	Yes	Yes	Yes	Yes
Control variables for household	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	24,126	24,126	24,126	24,126
R^2	0.130	0.083	0.130	0.105
Wald value		56.520		58.313
first stage F-value		88.11		91.55

****Indicate significant at the 1% statistical levels, respectively, with t-values in parentheses.*

TABLE 4 | Robustness test: replacement of explanatory variables.

	(1)	(2)	(3)	(4)
Digital payments (0–1)	0.265***	0.218**	0.216**	0.194**
	(6.56)	(2.68)	(2.57)	(2.49)
Control variables for household head	No	No	Yes	Yes
Control variables for household	No	No	No	Yes
Household fixed effects	No	Yes	Yes	Yes
Time fixed effects	No	Yes	Yes	Yes
N	24,126	24,126	24,126	24,126
R^2	0.001	0.004	0.005	0.016

*** and ** indicate significant at the 1 and 5% statistical levels, respectively, with t-values in parentheses.*

there is no problem of weak instrumental variables and the instrumental variables we selected are valid. In addition, driving distance is an instrumental variable similar to spherical distance. Specifically, we use the driving distance from the capital city of the household's province to Hangzhou as the instrumental variable, and the results are still robust.

Replacing Core Explanatory Variables

Considering that the frequency of household heads using the Internet for commercial activities is not linear, in the robustness analysis, we choose to use whether household heads conduct commercial activities via the Internet as a proxy variable for digital payments in order to eliminate the effect of uneven frequency distribution. If the household head usually does not ever conduct business activities on the Internet, then we consider that the household head has not used digital payments and the core explanatory variable is equal to 0. Conversely, if the household head usually conducts business activities on the Internet, then we consider that the household head has used digital payments and the core explanatory variable is equal to 1.

TABLE 5 | Robustness tests: replacement of explanatory variables.

	(1)	(2)	(3)	(4)
	Medical expenditure	**Fitness expenditure**	**Median (0–1)**	**Average (0–1)**
Digital payments	0.0432**	0.157***	0.00840**	0.00790**
	(2.20)	(8.71)	(2.28)	(2.63)
Control variables for household head	Yes	Yes	Yes	Yes
Control variables for household	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
N	24,126	24,126	24,126	24,126
R^2	0.782	0.098	0.018	0.017

*** and ** indicate significant at the 1 and 5% statistical levels, respectively, with t-values in parentheses.*

We know from **Table 4** that the use of digital payments by households promotes household healthcare expenditure.

Replacing the Explanatory Variables

Household healthcare expenditure can be divided into two components: household medical expenditure and household fitness expenditure, and both components can represent to some extent the degree to which households value their health. To conduct robustness tests, we split the household healthcare expenditure into two sub-dimensions, medical expenditure and fitness expenditure. In addition, we also replace the explanatory variables with dummy variables based on the median and mean of household healthcare expenditure. If the household healthcare expenditure exceeds the median or the mean, the explained variable equals 1. As seen in **Table 5**, the results are still robust.

TABLE 6 | Robustness test: replacement sample range.

	(1)	(2)	(3)
	Winsorize 1%	**Winsorize 5%**	**Winsorize 10%**
Digital payments	0.0559**	0.0534**	0.0534**
	(2.17)	(2.07)	(2.07)
Control variables for household head	Yes	Yes	Yes
Control variables for household	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	24,126	24,126	24,126
R^2	0.016	0.016	0.016

***Indicate significant at the 5% statistical levels, respectively, with t-values in parentheses.*

Replacing the Sample Range

To prevent some extreme values and outliers from affecting the study, we next apply winsorizing to the sample data to correct for all variables and address possible measurement errors in the regression. We can see from the results in columns (1) to (3) in **Table 6** that the use of big data-based digital payments by households increases household healthcare expenditure after a 1, 5, and 10% winsorizing process, respectively.

Mechanism Analysis

It can be found from the previous section that the use of big data-based digital payments by households significantly contributes to household healthcare expenditure. We further discuss and analyze the mechanism by which digital payments affect household healthcare expenditure using a mediating effects model, which is set up as follows.

$$Expenditure_{ijt} = \alpha_1 + \beta_1 Digital\ payments_{ijt} + \gamma_1 Control_{ijt} + \varphi_i + \delta_t + \varepsilon_{ijt} \quad (2)$$

$$Inter_{ijt} = \alpha_2 + \beta_2 Digital\ payments_{ijt} + \gamma_2 Control_{ijt} + \varphi_i + \delta_t + \varepsilon_{ijt} \quad (3)$$

$$Expenditure_{ijt} = \alpha_3 + \beta_3 Digital\ payments_{ijt} + \theta Inter_{ijt} + \gamma_3 Control_{ijt} + \varphi_i + \delta_t + \varepsilon_{ijt} \quad (4)$$

The *Inter* are the mediating variables, and the *Control* is the control variable, which is consistent with the baseline regression. The core coefficient of the mediation model is β, if the coefficients in models (2) and (3) are significant, and pass the mediating effect test and are consistent with the theoretical hypothesis, then the mediating effect exists.

Household Use of Big Data-Based Digital Payments Eases Household Credit Constraints

We use the household's loan balance to determine credit availability and thus measure the household's credit constraint status. If the household has any debt, then credit availability is equal to 1. If the household has no debt, then its credit availability is equal to 0. In the CFPS questionnaire, the relevant questions related to credit availability are: "Does your family have any outstanding bank loans to buy or build or renovate your house?", "Apart from your mortgage, does your family have any other outstanding bank loans at present?" "Do you have any outstanding loans from organizations or individuals other than banks, such as private credit institutions, relatives, friends, acquaintances, etc., for the purchase or construction or renovation of your home?" and "Pending loans from friends, relatives and private individuals" Does your household owe money to organizations or individuals other than banks (e.g., private credit institutions, relatives, friends, acquaintances, etc.) for reasons other than borrowing for the purchase or construction of a house, and has it not been repaid?" If the household head answers "yes" to any of the above questions, the household is certified as having a loan month and their credit availability is 1.

TABLE 7 | Mechanism analysis: credit constraints.

	(1)	(2)	(3)
Dependent variable	**Household healthcare expenditure**	**Credit Availability**	**Household healthcare expenditure**
Digital payments	0.0534**	0.00892***	0.0506*
	(2.07)	(2.86)	(1.95)
Credit Availability			0.315***
			(5.87)
Control variables for household head	Yes	Yes	Yes
Control variables for household	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	24,126	24,126	24,126
R^2	0.016	0.006	0.018
Sobel Z	2.803		
Sobel Z-P values	(0.0005)		
Bootstrap Z	2.60		
Bootstrap Z-P values	(0.009)		
Intermediary effect as a percentage	2.30%		

****, **, and * indicate significant at the 1, 5, and 10% statistical levels, respectively, with t-values in parentheses.*

As we can see from column (2) in **Table 7**, the use of digital payments by households has a significant positive effect on the credit availability of households. We can also see by column (3) that the coefficient of the core explanatory variable becomes smaller when credit availability is added to the regression. Moreover, both the Sobel test and the Boostrap test show that the mediating effect is significant at the 1% confidence level, a phenomenon that confirms the existence of a partial mediating effect and suggests that the mediating effect explains 2.3% of the total effect.

The reason is that households generate big data when they use digital payments, and these big data can empower both

the supply side and demand measurement of financial services, reducing households' credit constraints and ultimately benefiting the increase of households' healthcare spending. Specifically, on the one hand, financial institutions use big data to analyze the payment behavior activities of households, which helps financial institutions accurately assess the risk status of households, thus the risk of financial institutions providing credit to households is reduced, and financial institutions have more incentives to provide loans to households. On the other hand, big data can provide households with precise information of various types, including not only information about credit services that are suitable for them, but also information about the household's financial income and employment. As a result, households are more likely to increase their applications for credit services, and their incomes are likely to increase. And higher-income households are more likely to have their credit applications approved. In summary, household use of big data-based digital payments promotes household healthcare spending by easing household credit constraints.

Household Use of Big Data-Based Digital Payments Increases the Social Capital of Households

We use household spending on human gifts to measure the social capital of the household. With respect to favor gift expenditure, the CFPS questionnaire addresses the question, "Including in-kind and cash, how many total favor gifts did your household give in the past 12 months?" In order to correct for bias, we talk about the expenditure on favors to be logarithmic. The specific regression results can be seen in **Table 8**. We can see through column (2) that the use of digital payments by households has a significant positive effect on social capital. We can also see through column (3) that the coefficients of the core explanatory variables become smaller when social capital is added to the regression. In addition, both the Sobel test and the Boostrap test show that the mediation effect is significant at the 1% confidence level, a phenomenon that confirms the existence of a partial mediation effect and suggests that the mediation effect explains 7.24% of the total effect.

In other words, when households use digital payments based on big data, digital platforms exert social attributes. The digital platform uses big data to mine families' usage data and promotes subjects on the digital platform to families, such as relatives, friends, and medical institutions, in order to strengthen their dependency on the digital platform while also increasing their social capital. On the one hand, increasing a family's social capital allows them to share knowledge with other people and expand their information channels. Families may be able to receive more healthcare information via their social networks, lowering the cost of looking for healthcare information and allowing households to increase healthcare spending. On the other hand, increasing households' social capital allows them to be more closely connected with medical institutions, banks, and other subjects, allowing medical institutions and banks to better understand households' situations and provide health care services to them, thereby increasing household health care expenditures.

TABLE 8 | Mechanism analysis: social capital.

	(1)	(2)	(3)
Dependent variable	**Household healthcare expenditure**	**Social Capital**	**Household healthcare expenditure**
Digital payments	0.0525*	0.0344*	0.0493*
	(1.93)	(1.86)	(1.84)
Social Capital			0.0951***
			(8.02)
Control variables for household head	Yes	Yes	Yes
Control variables for household	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
N	24,126	24,126	24,126
R2	0.016	0.046	0.025
Sobel Z	5.543		
Sobel Z-P values	(0.000)		
Bootstrap Z	5.880		
Bootstrap Z-P values	(0.000)		
Intermediary effect as a percentage	7.24%		

**** and * indicate significant at the 1 and 10% statistical levels, respectively, with t-values in parentheses.*

TABLE 9 | Heterogeneity of household financial accessibility.

	(1)	(2)
	Low financial accessibility	**High financial accessibility**
Digital payments	0.0294	0.0809**
	(0.89)	(2.29)
Control variables for household head	Yes	Yes
Control variables for household	Yes	Yes
Household fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
N	14,092	10,034
R^2	0.017	0.017

***Indicate significant at the 5% statistical levels, respectively, with t-values in parentheses.*

FURTHER DISCUSSION: THE HETEROGENEITY OF HOUSEHOLD HEALTHCARE EXPENDITURES

The previous analysis found that households' digital payment use increases households' healthcare spending by alleviating credit and information constraints. So, does this effect differ for households with different endowments? We examine the heterogeneous impact of digital payment use on households' health care expenditures separately from differences in financial

TABLE 10 | Heterogeneity in the age of household heads.

	(1)	(2)
	Young and middle-aged people (18–59 years old)	**Elderly (60 years old and above)**
Digital payments	0.0563*	0.0766
	(1.99)	(1.39)
Control variables for household head	Yes	Yes
Control variables for household	Yes	Yes
Household fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
N	16,142	7,984
R^2	0.019	0.015

**Indicate significant at the 10% statistical levels, respectively, with t-values in parentheses.*

TABLE 11 | Heterogeneity in education level of household heads.

	(1)	(2)
	Lower level of education	**Higher level of education**
Digital payments	−0.0454	0.0536*
	(-0.59)	(1.98)
Control variables for household head	Yes	Yes
Control variables for household	Yes	Yes
Household fixed effects	Yes	Yes
Time fixed effects	Yes	Yes
N	6,050	18,076
R^2	0.009	0.020

**Indicate significant at the 10% statistical levels, respectively, with t-values in parentheses.*

availability, age of the household head, and education level of the household head.

Financial Accessibility of Households

In this paper, we refer to Beck et al. (27) and use the number of bank branches per capita in each province as a proxy variable for traditional financial accessibility. After that, this paper divides the sample into two groups, low traditional financial accessibility and high traditional financial accessibility, based on the average of the number of bank branches per capita in each province. As seen in **Table 9**, the coefficient of digital payment in areas with low traditional financial accessibility is 0.0809 and is significant at the 5% level. This indicates that in places with high traditional financial accessibility, households can more easily participate in traditional financial services and have higher financial literacy. Therefore, this group can use digital payments more easily and give full play to the role of digital payments empowered by big data.

The Age of the Household Head

The household head has a substantial impact on the household's decision-making, and age is an essential aspect of the household head. Based on whether the household head is <60 years old, we split the sample into a young and middle-aged group and an older group. Young and middle-aged household heads usually have a higher level of awareness, recognize the advantages of digital payments, and have a stronger willingness to use them. In addition, young and middle-aged heads of households usually have a higher risk tolerance. They are also more open-minded, and they are more willing to try new things. Most older heads of household may have a relatively lower level of awareness and are more traditional and conservative in their thinking. They will be more concerned about the risks associated with digital payments, such as suffering losses due to fraud, than the benefits that come with them. So they may be reluctant to use digital payments.

As seen in **Table 10**, in the young and middle-aged group, the coefficient of digital payments is 0.0563 and is significant at the 10% confidence level. This result is also more consistent with our common sense. Generally, younger household heads are more capable of learning, and they are receptive and willing to learn to use digital payments.

Education Level of the Household Head

We separated the sample into two groups, less educated and more educated, depending on whether the household head has fewer than 6 years of school, and then regressed them, because around 43.37 percent of household heads had <6 years of education. The results in **Table 11** show that the coefficient of digital payments in the group with higher education level is 0.0536 and is significant at the 10% confidence level. These results suggest that the more educated household heads may have a stronger learning ability and are more receptive to new things such as digital payments, so they are able to take full advantage of digital payments and thus increase their healthcare spending.

CONCLUSION

This article investigates the effect of big data-based digital payments on household healthcare expenditure from a big-data perspective. This research analyzes 24,126 sample data from 2014 to 2018 from the China Family Panel Studies (CFPS) to investigate the effect of big data-based digital payments on household healthcare expenditure. This research uses benchmark regression to examine the intermediary effect and heterogeneous influence of household digital payment on household health care expenditure, in order to improve knowledge of household health care expenditure and assist families in making better use of digital payment and maximizing its role in promoting a healthy lifestyle.

The following are the findings of this study. First, big data-based digital payments can dramatically raise household healthcare expenditure; that is, the higher a family's mobile spending, the higher the household healthcare expenditure. Second, the mechanism analysis demonstrates that by reducing household credit limitations and collecting social capital, the household application of big data-based digital payments can enhance household health care expenditure. Third, in families

headed by middle-aged and young individuals with high education and financial access, the use of digital payment has a larger role in promoting household healthcare expenditure. Fourth, the conclusions of this work are still robust when using the instrumental variable technique, substituting the indicators of household healthcare expenditure, replacing the indicators of digital payment, and shortening the tail of the variables.

This paper, however, may have the following flaws. First, this article only looks at the influence of digital payments on household healthcare expenditure, but it ignores the impact on individual health outcomes, which is a finding that needs to be explored more in the future. Second, because China's digital payment includes a variety of types, including mobile payment, bank electronic payment, digital RMB payment, etc., this paper examines the impact of digital payment in general from the perspective of big data, rather than providing a more detailed classified discussion on digital payment. As a result, researching the effects of various digital payment methods is an area that has to be addressed in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CL and DL contributed to conception and design of the study. SH organized the database. SS performed the statistical analysis. YT wrote the first draft of the manuscript. CL, DL, and ZW wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

1. Rufai AM, Ogunniyi AI, Abioye OD, Birindwa AB, Olagunju KO, Omotayo AO. Does economic shocks influence household's healthcare expenditure? Evidence from rural Nigeria. *Heliyon.* (2021) 7:e06897. doi: 10.1016/j.heliyon.2021.e06897
2. Sultana M, Mahumud RA, Sarker AR. Burden of chronic illness and associated disabilities in Bangladesh: evidence from the Household Income and Expenditure Survey. *Chronic Dis Transl Med.* (2017) 3:112–22. doi: 10.1016/j.cdtm.2017.02.001
3. Mahumud RA, Sarker AR, Sultana M, Sheikh N, Islam M, Hossain M. *The Determinants Out-of-Pocket Healthcare Expenditure in Bangladesh: Evidence from Household Income and Expenditure Survey-2010[M]//Issues on Health and Healthcare in India*. Singapore: Springer (2018). p. 339–50.
4. Cader I, Dilrukshi M, Senanayake K, Welgama M. Household income and healthcare expenditure in Sri Lanka. In: *Proceedings of the Vavuniya Campus international Research Symposium.* (2019).
5. Lee YJ, Lee HO. Influence and change of healthcare expenditure on household income inequality. *J Korea Contents Assoc.* (2019) 19:331–41. doi: 10.5392/JKCA.2019.19.05.331
6. Molla A, Chi C. How much household healthcare expenditure contributes to poverty? Evidence from the Bangladesh household income and expenditure survey, 2010. *J Poverty.* (2020) 24:627–41. doi: 10.1080/10875549.2020.1742269
7. Sarker AR, Mahumud RA, Sultana M, Ahmed S, Ahmed W, Khan JA. The impact of age and sex on healthcare expenditure of households in Bangladesh. *Springerplus.* (2014) 3:1–5. doi: 10.1186/2193-1801-3-435
8. Maharana B, Ladusingh L. How does the change in household age-sex composition affect out of pocket healthcare expenditure of older adults in India? *Ageing Int.* (2021) 1–22. doi: 10.1007/s12126-021-09457-3
9. Mahmud KT, Hilton D. Does microcredit really matter for healthcare expenditure of the poor fish-farmers? *Perspective from rural Bangladesh. Journal of Poverty.* (2020) 24:147–67. doi: 10.1080/10875549.2019.1668901
10. Levchenko AA. Financial liberalization and consumption volatility in developing countries. *IMFStaff Pap.* (2005) 52:237–59. doi: 10.5089/9781589064485.024
11. Taylor E. Mobile payment technologies in retail: a review of potential benefits and risks. *Int J Retail Distribution Manag.* (2016) 44:159–177. doi: 10.1108/IJRDM-05-2015-0065
12. Zhao C, Wu Y, Guo J. Mobile payment and Chinese rural household consumption. *China Econ Rev.* (2022) 71:101719. doi: 10.1016/j.chieco.2021.101719
13. Zhang J, Zhang H, Gong X. Mobile payment and rural household consumption: evidence from China. *Telecomm Policy.* (2022) 46:102276. doi: 10.1016/j.telpol.2021.102276
14. Ahmed H, Cowan B. Mobile money and healthcare use: Evidence from East Africa. *World Develop.* (2021) 141:105392. doi: 10.1016/j.worlddev.2021.105392
15. Song Q, Li J, Wu Y, Yin Z. Accessibility of financial services and household consumption in China: evidence from micro data. *N Am J Econ Finance.* (2020) 53:101213. doi: 10.1016/j.najef.2020.101213
16. Li C, Lin L, Gan CEC. China credit constraints and rural households' consumption expenditure. *Finance Res Lett.* (2016) 19:158–64. doi: 10.1016/j.frl.2016.07.007
17. Hou L, Hsueh SC, Zhang S. Digital payments and households' consumption: a mental accounting interpretation. *Emerg Markets Finance Trade.* (2021) 57:2079–93. doi: 10.1080/1540496X.2021.1887727
18. Kang X. The impact of family social network on household consumption. *Modern Econ.* (2019) 10:679. doi: 10.4236/me.2019.103046
19. Jack W, Suri T. Risk sharing and transactions costs: evidence from Kenya's mobile money revolution. *Am Econ Rev.* (2014) 104:183–223. doi: 10.1257/aer.104.1.183
20. Xiao JJ, Chen C, Chen F. Consumer financial capability and financial satisfaction. *Soc Indic Res.* (2014) 118:415–32. doi: 10.1007/s11205-013-0414-8
21. Fujiki H, Tanaka M. Currency demand, new technology, and the adoption of electronic money: micro evidence from Japan. *Econ Lett.* (2014) 125:5–8. doi: 10.1016/j.econlet.2014.07.032
22. Martikainen E, Schmiedel H, Takalo T. Convergence of European retail payments. *J Bank Finance.* (2015) 50:81–91. doi: 10.1016/j.jbankfin.2014.09.021
23. Li J, Wu Y, Xiao JJ. The impact of digital finance on household consumption: evidence from China. *Econ Model.* (2020) 86:317–26. doi: 10.1016/j.econmod.2019.09.027
24. Mackenzie A. Personalization and probabilities: impersonal propensities in online grocery shopping. *Big Data Soc.* (2018) 5:2053951718778310. doi: 10.1177/2053951718778310
25. Yildirim M, Okay FY. Özdemir S. Big data analytics for default prediction using graph theory. *Expert Syst Appl.* (2021) 176:114840. doi: 10.1016/j.eswa.2021.114840
26. Dupas P. What matters (and what does not) in households' decision to invest in malaria prevention?. *Am Econ Rev.* (2009) 99:224–30. doi: 10.1257/aer.99.2.224
27. Beck T, Pamuk H, Ramrattan R, Uras BR. Payment instruments, finance and development. *J Dev Econ.* (2018) 133:162–86. doi: 10.1016/j.jdeveco.2018.01.005

28. Wang X. Mobile payment and informal business: evidence from China's household panel data. *China World Econ.* (2020) 28:90–115. doi: 10.1111/cwe.12334
29. Pérez-Martín A, Pérez-Torregrosa A, Vaca M. Big Data techniques to measure credit banking risk in home equity loans. *J Bus Res.* (2018) 89:448–54. doi: 10.1016/j.jbusres.2018.02.008
30. Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning. *Rev Financ Stud.* (2020) 33:2223–73. doi: 10.1093/rfs/hhaa009
31. Sheng T. The effect of fintech on banks' credit provision to SMEs: evidence from China. *Finance Res Lett.* (2021) 39:101558. doi: 10.1016/j.frl.2020.101558
32. Lai JT, Yan IKM, Yi X, Zhang H. Digital financial inclusion and consumption smoothing in China. *China World Econ.* (2020) 28:64–93. doi: 10.1111/cwe.12312
33. Fungáčová Z, Weill L. Understanding financial inclusion in China. *China Econ Rev.* (2015) 34:196–206. doi: 10.1016/j.chieco.2014.12.004
34. Li R, Li Q, Huang S, Zhu X. The credit rationing of Chinese rural households and its welfare loss: an investigation based on panel data. *China Econ Rev.* (2013) 26:17–27. doi: 10.1016/j.chieco.2013.03.004
35. Yeung G, He C, Zhang P. Rural banking in China: geographically accessible but still financially excluded?. *Reg Stud.* (2017) 51:297–312. doi: 10.1080/00343404.2015.1100283
36. Bello-Orgaz G, Jung JJ, Camacho D. Social big data: recent achievements and new challenges. *Inform Fusion.* (2016) 28:45–59. doi: 10.1016/j.inffus.2015.08.005
37. Dorsey S, Forehand R. The relation of social capital to child psychosocial adjustment difficulties: the role of positive parenting and neighborhood. *J Psychopathol Behav Asses.* (2003) 1:11–23. doi: 10.1023/A:1022295802449
38. Hancock D, Humphrey DB, Wilcox JA. Cost reductions in electronic payments: the roles of consolidation, economies of scale, technical change. *J Bank Finance.* (1999) 23:391–421. doi: 10.1016/S0378-4266(98)00108-3
39. Humphrey DB, Pulley LB, Vesala JM. Cash, paper, and electronic payments: a cross-country analysis. *J Money Credit Bank.* (1996) 28:914–39. doi: 10.2307/2077928
40. Shakhovska N, Fedushko S, Melnykova N, Shvorob I, Syerov Y. Big Data analysis in development of personalized medical system. *Procedia Comput Sci.* (2019) 160:229–34. doi: 10.1016/j.procs.2019.09.461
41. Björkegren D, Grissen D. The potential of digital credit to bank the poor. *Am Econ Rev.* (2018) 108:68–71. doi: 10.1257/pandp.20181032
42. Lusardi A, Mitchell OS, Curto V. Financial literacy among the young. *J Consumer Affairs.* (2010) 44:358–80. doi: 10.1111/j.1745-6606.2010.01173.x
43. Heineck G, Anger S. The returns to cognitive abilities and personality traits in Germany. *Labour Econ.* (2010) 17:535–46. doi: 10.1016/j.labeco.2009.06.001
44. Lindqvist E, Vestman R. The labor market returns to cognitive and noncognitive ability: evidence from the Swedish enlistment. *Am Econ J Appl Econ.* (2011) 3:101–28. doi: 10.1257/app.3.1.101
45. Agarwal S, Amromin G, Ben-David I, Chomsisengphet S, Evanoff DD. Financial literacy and financial planning: evidence from India. *J Hous Econ.* (2015) 27:4–21. doi: 10.1016/j.jhe.2015.02.003
46. Nasri W, Charfeddine L. Factors affecting the adoption of Internet banking in Tunisia: an integration theory of acceptance model and theory of planned behavior. *J High Technol Manag Res.* (2012) 23:1–14. doi: 10.1016/j.hitech.2012.03.001
47. Leyshon A, Thrift N. The restructuring of the UK financial services industry in the 1990s: a reversal of fortune?. *J Rural Stud.* (1993) 9:223–41. doi: 10.1016/0743-0167(93)90068-U
48. Dymski GA, Veitch JM. Financial transformation and the metropolis: booms, busts, and banking in Los Angeles. *Environ Plann A.* (1996) 28:1233–60. doi: 10.1068/a281233
49. Liu T, Pan B, Yin Z. Pandemic, mobile payment, and household consumption: micro-evidence from China. *Emerg Markets Finance Trade.* (2020) 56:2378–89. doi: 10.1080/1540496X.2020.1788539
50. Zhu K, Zhou Y, Zhao J. Guaxi and financial exclusion: empirical evidence from households in China. *Pacific Basin Finance J.* (2021) 67:101566. doi: 10.1016/j.pacfin.2021.101566
51. Wang J, Ai S, Huang M. Migration history, hukou status, and urban household consumption. *Econ Model.* (2021) 97:437–48. doi: 10.1016/j.econmod.2020.04.012
52. Zhou L, Yang JJ. Air pollution and household medical expenses: evidence from China. *Front Public Health.* (2021) 9:798780. doi: 10.3389/fpubh.2021.798780
53. Mekonen AM, Gebregziabher MG, Teferra AS. The effect of community based health insurance on catastrophic health expenditure in Northeast Ethiopia: a cross sectional study. *PLoS ONE.* (2018) 13:e0205972. doi: 10.1371/journal.pone.0205972
54. Mohanty SK, Chauhan RK, Mazumdar S, Srivastava A. Out-of-pocket expenditure on health care among elderly and non-elderly households in India. *Soc Indic Res.* (2014) 115:1137–57. doi: 10.1007/s11205-013-0261-7
55. Sekhampu TJ, Niyimbanira F. Analysis of the factors influencing household expenditure in a South African township. *Int Bus Econom Res J.* (2013) 12:279–84. doi: 10.19030/iber.v12i3.7671
56. Zhang Z, Ma C, Wang A. A longitudinal study of multidimensional poverty in rural China from 2010 to 2018. *Econ Lett.* (2021) 204:109912. doi: 10.1016/j.econlet.2021.109912

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Li, He, Sun, Tian and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fpubh.2022.846598

Research on Infant Health Diagnosis and Intelligence Development Based on Machine Learning and Health Information Statistics

Siyu Wang[1]*, *Min Li*[1] and *Soo Boon Ng*[2]

[1] Teachers College, Chengdu University, Chengdu, China, [2] SEGI University, Petaling Jaya, Malaysia

Intelligent health diagnosis for young children aims at maintaining and promoting the healthy development of young children, aiming to make young children have a healthy state and provide a better future for their physical and mental health development. The biological basis of intelligence is the structure and function of human brain and the key to improve the intelligence level of infants is to improve the quality of brain development, especially the early development of brain. Based on machine learning and health information statistics, this paper studies the development of infant health diagnosis and intelligence, physical and mental health. Pre-process the sample data, and use the filtering method based on machine learning and health information statistics for feature screening. Compared with traditional statistical methods, machine learning and health information statistical methods can better obtain the hidden information in the big data of children's physical and mental health development, and have better learning ability and generalization ability. The machine learning theory is used to analyze and mine the infant's health diagnosis and intelligence development, establish a health state model, and intuitively show people the health status of their infant's physical and mental health development by means of data. Moreover, the accumulation of these big data is very important in the field of medical and health research driven by big data.

Keywords: machine learning, health information statistics, infant health diagnosis, intelligence development, big data

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Abdul Rehman Javed, Air University, Pakistan
Praveen Kumar, VIT University, India

***Correspondence:**
Siyu Wang
wangsiyu@cdu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *31 December 2021*
Accepted: *22 February 2022*
Published: *02 June 2022*

Citation:
Wang S, Li M and Ng SB (2022) Research on Infant Health Diagnosis and Intelligence Development Based on Machine Learning and Health Information Statistics. Front. Public Health 10:846598. doi: 10.3389/fpubh.2022.846598

INTRODUCTION

In the process of improving people's quality of life, China pays more and more attention to the healthy development of human beings. Although people have deepened their understanding of human health under the promotion of this information age, China's research in the field of children's physical and mental health development still needs to be further deepened (1). Intelligence, also known as intelligence, refers to the people's response and processing ability to the surrounding environment and things. Human beings have extremely high intelligence, which is characterized by observation ability, memory ability, and thinking ability. Children's health diagnosis and intellectual development are extremely important factors during their childhood and even in their future life. Only when children grow up healthily, they will get good development in the future (2). In the field of children's health, diagnosis of children's intelligent health is the top priority. Children's intelligent health diagnosis aims to maintain and promote the children's

healthy development, and is determined to let children have a good state of health and provide a better future for the children's physical and mental health development (3). The biological basis of intelligence is the structure and function of human brain. The key to improve the intelligence level of infants is to improve the quality of brain development, especially the early development of brain. Children's physical and mental health development inspectors are still in the manual stage of processing the children's physical condition data, which not only increases the workload of children's physical condition inspectors, but also may delay the treatment of children's condition (4). In the great project of improving population quality and cultivating socialist successors, the cultivation of infants' health diagnosis and intellectual development ability is a key link that cannot be ignored. It is the basic project to improve the children's physical and mental health development. This foundation must be laid well and firmly. With the rapid progress and development of information technology, various industries have accumulated a large amount of information and data. How to mine valuable information from these data is a problem that all fields need to face (5). As one of the main methods of data analysis, machine learning is widely used in data mining.

Based on machine learning and health information statistics, this paper studies the development of infant health diagnosis and intelligence, physical and mental health. Pre-process the sample data and use the filtering method based on machine learning and health information statistics for feature screening. Finally, using multiple machine learning algorithms, from single classifier, integrated classifier to fusion classifier, the advantages and disadvantages of multiple machine learning models are systematically explored to improve the prediction accuracy and fitting effect of the models (6, 7). Compared with traditional statistical methods, machine learning and health information statistical methods can better obtain the hidden information in the big data of children's physical and mental health development, and have a better learning ability and generalization ability (8). Machine learning is the study of making computers to imitate the way of human learning and thinking, and making computers have the ability of self-learning. Through learning, the computer can constantly update and reorganize the existing cognitive style, so that its performance can be continuously improved. Machine learning is the core of the development of artificial intelligence and the fundamental way to make inanimate computers intelligent. Machine learning is a process in which computers can identify and acquire knowledge through programming language and mathematical principles, and continuously improve the performance of infant health diagnosis and intelligence development, which can be roughly divided into supervised learning and unsupervised learning. The self-learning ability of machine learning is very important for data analysis, and it is of great significance for regression prediction and classification problems. The regression function can play a great role in finance, economic market, transportation, and benefit management. The core idea of machine learning is not to design a specific learning mode to let the machine learning algorithm do it, but to let the machine learning algorithm learn the laws and essential characteristics of data by itself. There are two kinds of unsupervised learning ideas. The first unsupervised learning idea is that when the training network parameters don't specify accurate data set classification, it adopts a weighted incentive and punishment system, which can increase and decrease the size of network parameters in the network. Supervised learning establishes a prediction model by analyzing the logical relationship between features and target variables, so as to obtain the prediction results of unknown data (9). Unsupervised learning analyzes the logical relationship of sample data without objective variables. Clustering analysis and association rule analysis are typical unsupervised learning.

When there is category imbalance in the development of physical and mental health in the health information statistical dataset of machine learning, it will have a great impact on the classification sensitivity of machine learning (10). In fact, in the real world, the data are unbalanced. Class imbalance problems are widely used in many aspects. In the use of all descriptions, a few classes have high sensitivity, and its information is the target. In the big data technology based on machine learning and health information statistics; massive data sources are a necessary requirement for big data analysis (11). Based on the machine learning and health information statistics, newborn should be trained to crawl and walk on time, on the basis of child's health diagnosis of hand movements. Training child to crawl can expand the range of child's activities, enable them to move freely, and make efforts to get in touch with interested objects, which have a positive effect on the development of perception and intelligence. The healthcare for toddler around 1 year old is mainly to gradually train to stand and walk. Only with the data analysis method supported by sufficient data can we provide a good personalized solution for children's physical and mental health development for children's health diagnosis and intellectual development, and improve their living habits and lifestyle (12). When training infant language through machine learning, we should combine language with knowing concrete things. It is necessary to combine language teaching with specific activities, such as pointing to apples and saying "This is an apple" when giving children apples, and knowing the specific thing of apples while teaching the pronunciation of apples. Use machine learning theory to analyze and mine children's health diagnosis and intelligence development, establish a health state model, and intuitively show people the health status of children's physical and mental health development in the form of data (13, 14). Moreover, the continuous accumulation of these big data is very important in the scientific research in the field of medical and health driven by big data.

RELATED WORK

The study (15) suggests that early childhood is the top priority of all aspects of children's development, and the problems related to children's education during this period can't be ignored, especially the significance to children's development can't be underestimated. In the game, let children learn something, which has a profound impact on the development of children's intelligence. Early childhood is the most important period to

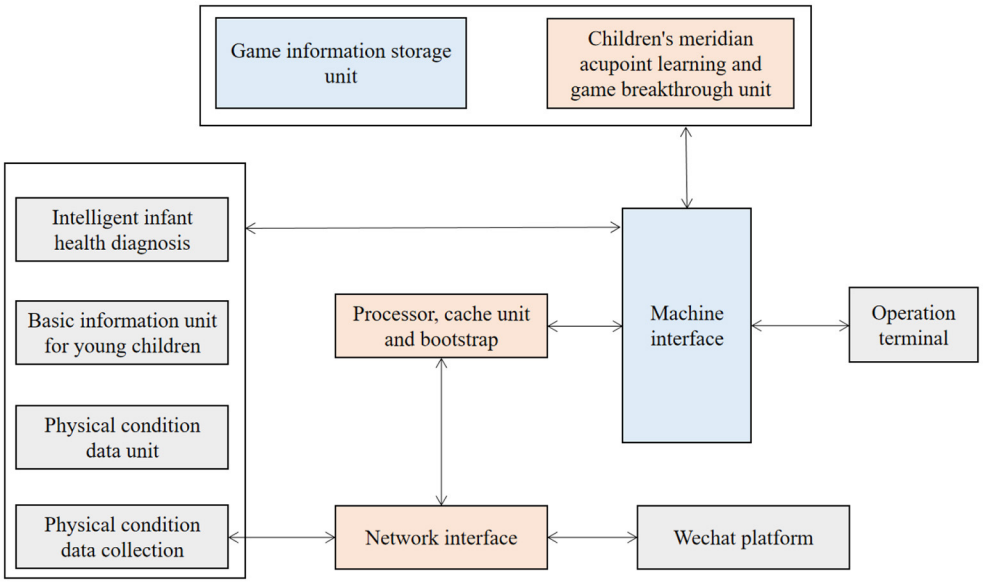


FIGURE 1 | Schematic diagram of the structure of infant intelligent health diagnosis and intelligence development.

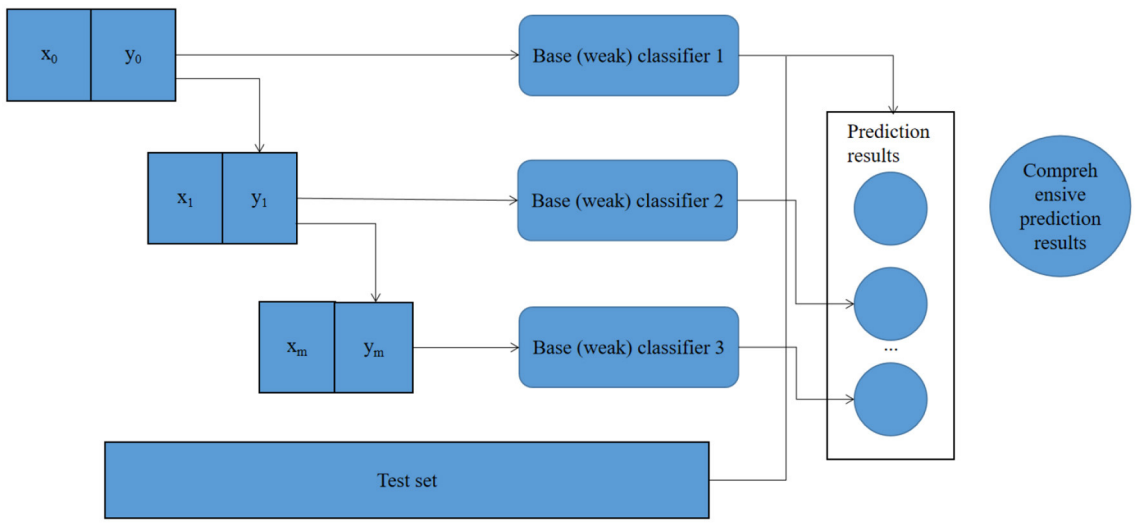


FIGURE 2 | Training flow chart of logistic regression algorithm model.

enhance the intelligence level of children. Education in the game mode is one of the important methods for enhancing the intelligence development of children.

The study (16) suggests that through the method of big data analysis, the newborn is crying, and others can't make it quiet, while the mother is in her arms. When the newborn touches the mother's skin and hears the mother's heartbeat, it will be quiet. Parents and relatives often make intimate and affectionate skin contact with the newborn, which is not only conducive to the development of the child's body and mind, but also conducive to the development of the child's intelligence.

A literature (17) research shows that 40–72% of the abnormal data processing actions of children's health data processing are an empty processing program or only record processing. In the worst case, this lack of proper action will not only not solve the problem, but even also lead to the bad development of the situation.

The study (18) suggests that mothers show affectionate skin contact by hugging, kissing, and gently massaging newborn child, especially skin contact and early sucking within half an hour after birth. This can not only compensate for the lost close skin contact in the mother's body, but also extend and expand the feeling of

comfort, so that the newborn can form a positive skin feeling conditioned reflex.

The study (19) shows that through the method of big data analysis, skin contact and massage is a physiological need of human beings and all warm-blooded animals. When child can't get this kind of need or satisfaction, skin hunger will happen, which will affect their psychological mood, normal psychological development, and intelligence development.

The study (20) shows that to improve the children's intellectual development, parents and teachers both need to drive children to participate in games, especially more games conducive to improving their intelligence, which cannot only make children have game experience, but also improve the children's subjective initiative and cognitive ability while playing games, so as to obtain intellectual development.

A study (21) suggests that the treatment of children's abnormal physical condition in kindergartens at this stage is still in the artificial stage, which is inefficient and takes a long time, which not only increases the workload of kindergarten staff, but also may delay the treatment of children's condition. The study (19) shows that through the big data analysis method, for kindergartens, today's preschool teachers choose to educate children. The contents that preschool teachers think are "interesting" and children are not easy to operate. In addition, they will place too much emphasis on learning in the classroom and ignore the diversity of learning styles. Therefore, this will not necessarily promote the development of children's intelligence.

Research (22) shows that after birth, because newborns love their mothers, when they come into contact with their mothers' skin, they will feel happy and have enough sense of security, which is also an important test for early newborn health care. Skin sensation is the earliest sensation in infants, followed by smell, taste, and hearing.

The study (23) proposes that children's information is uploaded to the cloud to update the database, and parents can view the children's information through mobile app. In 2017, Li Qingyi and song Xueliang disclosed an all-in-one kindergarten morning check-up and medical care machine, which proposed to integrate the kindergarten common disease check-up and nursing materials. Parents can view nursing suggestions through mobile app.

For the problem of infant health diagnosis and infant intelligence development, this research project proposes to study infant health diagnosis and intelligence development based on machine learning and health information statistics. Developing children's mental and physical health through learning and playing can improve the children's subjective initiative and help children's motivation, knowledge, and ability.

TABLE 1 | The comparison of the performance of data after three classification models applied to health information statistics for children's health diagnosis and intelligence development.

Model	Recall	Accuracy	F-score	A Accuracy	AUC
Logistic	0.602	0.935	0.733	0.813	0.895
Random forest	0.735	0.907	0.812	0.855	0.943
SVM	0.735	0.889	0.804	0.848	0.896

TABLE 2 | The performance comparison of data processed by health diagnosis and intelligence development of children with three classification models applied to health information statistics.

Model	Recall	Accuracy	F-score	Accuracy	AUC
Logistic	0.605	0.931	0.730	0.816	0.892
Random forest	0.738	0.913	0.815	0.857	0.947
SVM	0.736	0.892	0.814	0.847	0.899

TABLE 3 | Comparison of health diagnosis and intelligence development of children in kindergarten a.

Factor	Number of people surveyed	Number of people	Rate (%)	X2 value	*P* value
Small class	156	58	37	1.508	0.221
Middle shift	168	67	40	3.36	0.065
Taipan	197	89	45	4.571	0.064

TABLE 4 | Logistic regression model evaluation of basic data and combined data.

Input data	Data set	Health information statistics				
		Accuracy	Accuracy	Recall	F1 value	AUC
Basic data	Training set	74.39	75.78	73.47	74.60	82.24
	Test set	74.41	75.05	73.05	74.04	81.33
Combined data	Training set	74.91	75.60	75.28	75.45	84.18
	Test set	77.32	78.62	74.98	76.75	83.66

PRINCIPLE AND ALGORITHM OF MACHINE LEARNING

This topic integrates the automatic detection of children's physical and mental health development, the automatic management of children's information data and children's intelligence development. It has high degree of automation, strong reliability, and diversified functions, which is conducive to children's health and intelligence development. As shown in **Figure 1**.

The infant intelligent health diagnosis and intelligence development device comprises an operation terminal, an infant health self-test device, a processor, an infant intelligence development trainer, a computer interface, and a network interface. The operation terminal is connected with the processor through the network interface, and both the infant health self-test device and the infant intelligence development trainer are connected with the processor through the interface. The processor includes a processor and a storage unit, and the infant

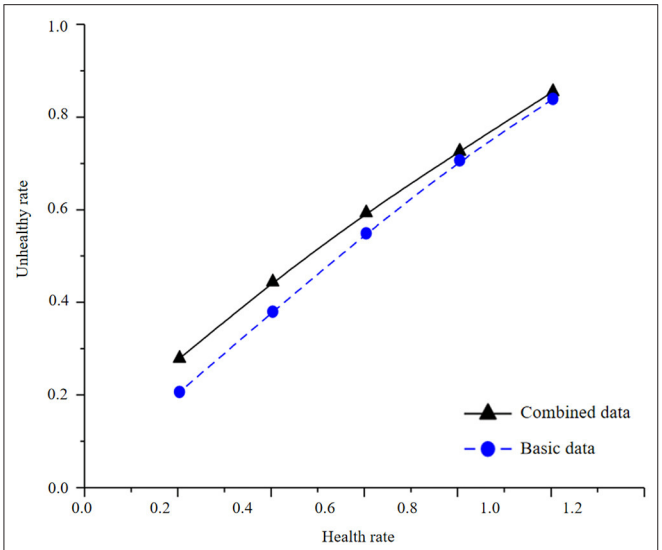


FIGURE 3 | Health information statistics of logistic regression combined data and basic data.

health self-test device and the infant intelligence development trainer, respectively, include the infant health self-test unit, the infant intelligence development training unit, and their respective boot programs and storage units.

Machine learning algorithm is used to construct the model of child health diagnosis and intelligence development. When using machine learning to solve problems, logistic regression is usually the first choice and then other complex algorithms are used on this basis. In order to solve the classification problem, firstly, the sigmoid function is used to construct the prediction function, that is, the classification task. Then, the $J(\theta)$ function is derived from the maximum likelihood estimation. Finally, the optimal model coefficients are solved by the gradient descent algorithm.

The function form is as follows:

$$g(z) = \frac{1}{1+e^{-z}} \tag{1}$$

The independent variable z is any real number, and the range is [0,1]. That is, the conversion from numerical value to probability is completed by sigmoid function, and the prediction function is as follows:

$$h_\theta(x) = g\left(\theta^T x\right) \frac{1}{1+e^{-\theta^t x}} \tag{2}$$

where x is the dataset introduced in the form of matrix, $\theta_0 x_0 + \theta_1 x_1 + \ldots\ldots + \theta_n x_n = \sum_{i=0}^{0} \theta_i x_i = \theta^T x$, and the best parameter is $\theta = [\theta_0, \theta_1, \ldots \theta_n]^T$.

For the binary classification task (0,1), the prediction function is constructed as follows:

$$P\left(y=1 \mid x,\theta\right) h_\theta(x), P\left(y=0 \mid x.\theta\right) = 1 - h_\theta(x) \tag{3}$$

For the sake of simplicity, the two-category tasks can also be integrated into

$$p\left(y \mid x;\theta\right) = \left(h_\theta(x)\right)^y \left(1-h_\theta(x)\right)^{1-y} \tag{4}$$

Based on the prediction function, the maximum likelihood estimation $J(\theta)$ function is constructed as follows:

$$J(\theta) = \frac{1}{m}\log p\left(y \mid x,\theta\right) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \log h_\theta(x_i) + (1-y_i)\log\left(1-h_\theta(x_i)\right)\right) \tag{5}$$

$J(\theta)$ converting the gradient ascending task into the gradient descending task.

The optimal coefficient θ_j^* can be obtained by solving the first-order equation of $J(\theta)$ to θ_j.

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x_i) - y_i\right) x_i^j \tag{6}$$

where x_i^j represents the sample value of row I and column J.

Finally, the optimal solution is obtained by parameter updating.

$$\theta_j^* = \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x_i) - y_i\right) x_i^j \tag{7}$$

The logistic regression requires simple input data of the model. The feature set can be classified variables or continuous variables, and the modeling process is also convenient. Therefore, it is a widely used classification model. In the field of disease prediction, unconditional logistic regression model is widely used.

Let the separation hyperplane separate the linearly separable sample set $T = \left\{\left(x_i, y_i\right), i = 1, \ldots, N\right\}$ correctly, that is

$$y_i = \begin{cases} 1 & w \cdot x_i + b \geq 1 \\ 0 & w \cdot x_i + b \leq 1 \end{cases} \tag{8}$$

The classification interval of the separation hyperplane is defined as

$$d = d_+ + d_- = \frac{|w \cdot x + b|}{\|w\|} = \frac{1}{\|w\|} \tag{9}$$

Among

$$d_+ = \min_{i:y=1}\left\{\left(w \cdot x_i + b\right) / \|w\|\right\} \tag{10}$$

$$d_- = \min_{i:y=-1}\left\{\left(w \cdot x_i + b\right) / \|w\|\right\} \tag{11}$$

Indicates the distance between the sample points and the decision surface equation. To maximize the hyperplane classification interval d, it can be transformed into a set of w &b to minimize $\|w\|$. For the later calculation aspect, it is transformed into a set of

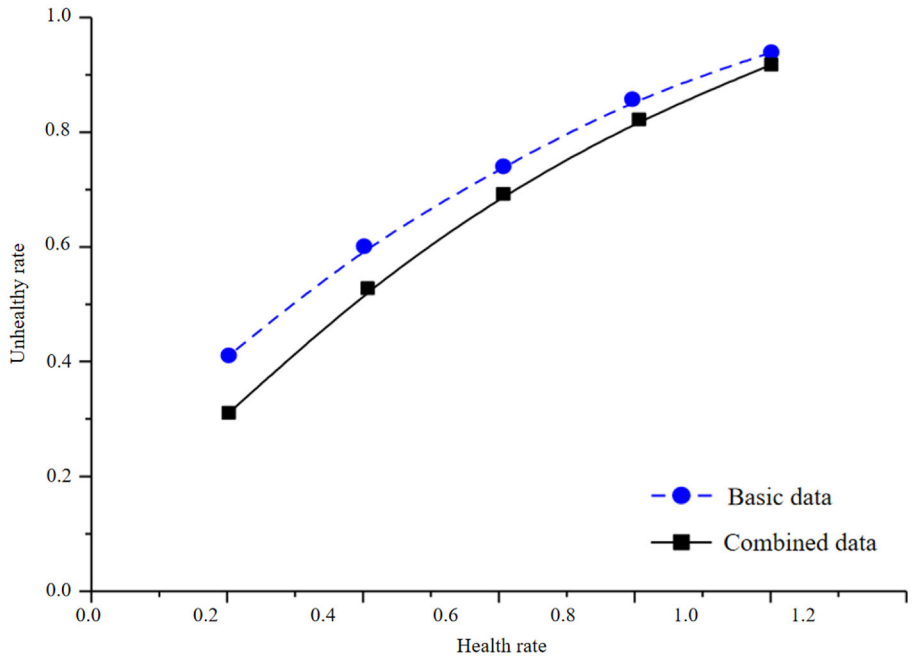


FIGURE 4 | Health information statistics of logistic regression combined data and basic data.

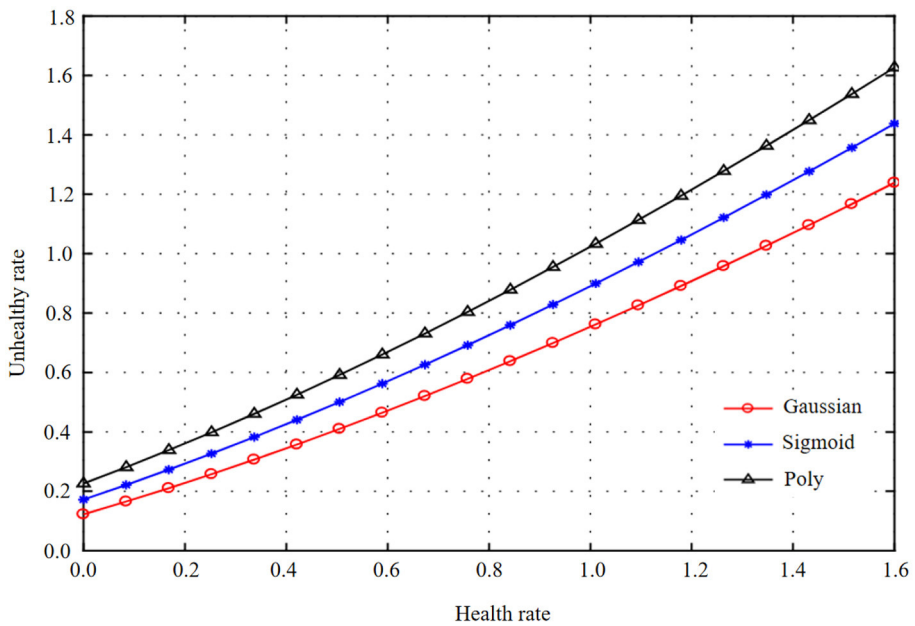


FIGURE 5 | Curve representations of three kernel functions of machine learning basic data and health information statistics.

w & b, which minimizes $\|w\|^2$. When the similarity between the collected characteristic information of infant's physical condition and a disease in the case information database is 0, the infant does not have the disease; when the similarity is not >0.4, the infant's physical condition is sub-healthy for the disease; when the similarity is >0.4, the infant is suspected to have the disease. Judgment of infant's physical health or suspected illness, when the similarity between the collected infant's physical condition characteristic information and all the illnesses in the case information base is 0, the infant is healthy; If the similarity of all diseases is not <0, i and not more than 0.4, the child's body is in a subhealth state; If the similarity of all diseases is not <0.4, the infant suffers from diseases, and the disease with higher similarity value is taken as the preliminary diagnosis result.

Different machine learning algorithms are obtained due to different function selection. However, the training process of

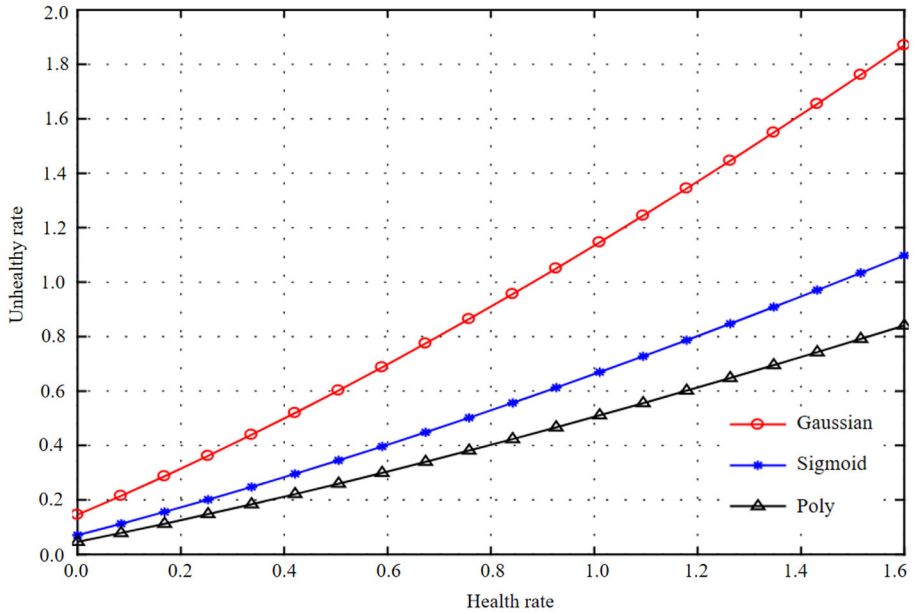


FIGURE 6 | Curve representations of three kernel functions of machine learning basic data and health information statistics.

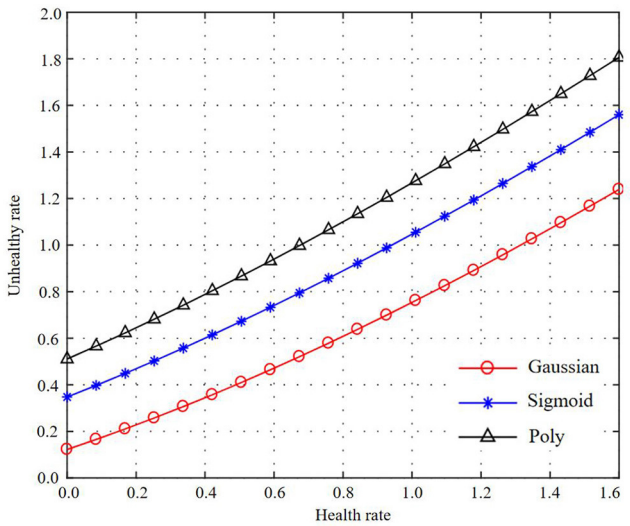


FIGURE 7 | Curve representation of three kernel functions of machine learning basic data and health information statistics.

logistic regression algorithm is to train each weak classifier in step form, and then combine the weak classifiers into strong classifiers in a certain way to obtain the comprehensive prediction results, as shown in **Figure 2**.

Considering the model principle and the data characteristics of this study, the logistic regression algorithm in machine learning algorithm will be used to construct the model of infant health diagnosis and intelligence development, and to find the most suitable model for infant health diagnosis and intelligence development.

INFANT HEALTH AND INTELLIGENCE DEVELOPMENT

Intelligent Diagnosis and Disease Treatment

Identification of Young Children

With the development of information technology, there are more and more forms of identification technology, which can be summarized into the following three categories: the first category is the things you can know, such as passwords, which are easy to crack and have the lowest relative security; The second category is the things you can own, such as smart cards, which are not easy to save, easy to lose, and low in security. The third category is the personal characteristics held by everyone, such as face and iris. These characteristics are unique to everyone. Compared with the first two categories, the third category is more trustworthy. Ensure the nutrition requirement of the first rapid proliferation of brain cells in the fetal period, and increase the number and quality of brain cells as much as possible with sufficient and high-quality nutrition. Pregnant women should pay attention to strengthening nutrition, eating reasonably and supplementing folic acid and trace elements such as zinc, iron, and copper. Iodine deficiency leads to dementia, and iodine deficiency areas should pay special attention to iodine supplementation. The important role of early education can't be replaced and compensated by any other form of education in the future. Long-term emotional stress and anxiety of pregnant women, fright, fear, sadness, high-dose radiation, organic mercury, lead and other toxic substances, tobacco and alcohol, virus infection, pregnancy poisoning, diabetes and hypothyroidism, taking phenytoin sodium, antithyroid drugs, perinatal hypoxia, and birth injury, etc. Grasp the "best age,"

TABLE 5 | Optimal parameter combination of different kernel functions modeling results of machine learning basic data and health information statistics.

Optimal parameter combination of different kernel functions	Data set	Health information statistics				
		Accuracy rate	Precision	Recall rate	F1 value	AUC
Ploy kernel function	Training set	72.78	86.87	55.17	67.45	84.18
C = 0.1, degree = 3 gamma = 0.01	Test set	71.63	83.72	53.63	65.37	80.45
Gaussian kernel function	Training set	76.63	78.72	74.48	76.54	84.28
C = 1, gamma = 0.01	Test set	76.23	77.68	73.53	75.55	82.57
Sigmoid kernel function	Training set	73.14	75.15	71.03	73.03	79.98
C = 0.1, gamma = 0.1	Test set	73.44	74.54	71.11	72.78	80.73

TABLE 6 | Modeling results of different kernel functions of SVM under machine learning combined data.

Optimal parameter combination of different kernel functions	Data set	Health information statistics				
		Accuracy rate	Precision	Recall rate	F1 value	AUC
Ploy kernel function	Training set	83.76	88.52	78.45	83.18	93.05
C = 0.1, degree = 3 gamma = 0.01	Test set	76.11	80.44	68.92	74.24	82.31
Gaussian kernel function	Training set	78.12	79.22	77.53	78.37	86.52
C = 1, gamma = 0.01	Test set	77.32	78.91	73.50	76.64	83.76
Sigmoid kernel function	Training set	73.40	78.91	64.83	71.21	81.53
C = 0.1, gamma = 0.1	Test set	75.38	78.44	66.51	72.96	82.45

that is, it is easier to learn certain knowledge or behavior at a certain age, faster to master and better to remember, and to carry out education in advance or delay, and the effect is far less than the best age. For babies, the facial movements should be trained first. Modern scientific research has proved that the more and more complex fingers move, the more they stimulate the brain, thus promoting the development of intelligence, so we should pay attention to the training of child's facial movements.

The specific steps of children's face training or recognition algorithm are as follows:

① Connect the camera.

② Camera device captures children's faces.

③ Extract and mark the captured children's faces.

④ For face training, name the captured child's face and store it in the child information database for training. For face recognition, compare the captured child's face with the child's face in the child information database. If the comparison is successful, extract the child's student status number, name, and other information; otherwise if you choose to continue, carry out infant face training. If you choose to end, carry out the next step.

⑤ End.

Children's Physical Condition Information Collection

Children's physical condition information collection is the premise of intelligent diagnosis of children's physical health. Its collection includes temperature acquisition, cough, diarrhea, runny nose, and so on. Genetically speaking, the development of human brain is influenced by genes and environment. The structure and function of brain are directly related to the quality of genes, but also influenced by environmental factors. Brain development is a process that determines the gradual expression of genes in the brain. In the process of gene expression, unfavorable or even harmful environmental factors can inhibit gene expression, causing brain development disorder or damage, while appropriate and favorable environmental factors can effectively promote gene expression, making the brain structure and function develop fully and perfectly. Early education is the most effective environmental stimulation factor to promote brain development, which can fundamentally improve the fine structure and function of brain, make it fully developed and perfect, not only improve the development quality of normal brain, but also significantly improve the intelligence level of mentally retarded people. Emphasize the nutritional conditions and further improve the quality of brain cells. Breast milk is the most ideal natural food, and breastfeeding is the most concerned issue in all countries today. The WHO requires that babies should be breastfed for at least 4 months after birth, just to meet the nutrition needs of the second rapid proliferation of brain cells. There are some differences in the reports about the influence of diet on physical development in different places. In some places, it is reported that children are eating too much animal protein at present; in some areas, children are eating too much plant food, and the proportion of nutrition intake among children in different areas is also different.

Strengthen the healthcare of infants' physical and mental health development. Infant immune system is imperfect and prone to diseases. Many diseases affect intellectual development. Therefore, active prevention and treatment should be carried out to minimize the occurrence of diseases and shorten the course of disease, so as to ensure the healthy development of the brain. In the past, when talking about the impact of diet on children's physical development, people immediately thought of malnutrition, which is indeed a very important aspect. How to ensure the healthy growth and development of children on the basis of solving the problem of malnutrition? There are many literature reports that with the improvement of material living standards, there are many new problems in children's physical development in China, such as children's obesity, anemia, precocity, physical decline, and so on. The non-contact infrared thermometer is used to obtain the temperature of children, which is wirelessly sent to the wireless receiver and stored in the collection unit of children's physical and mental health development. The utility model has the advantages of high accuracy of the measured body temperature. After obtaining the temperature of children, analyze whether the temperature is normal based on medical knowledge, and analyze the degree of fever for abnormal temperature. Biological studies have confirmed that the diet structure is closely related to children's physical and psychological growth and development, and is related to the occurrence of many diseases. In this paper, based on machine learning method, children's identity is recognized by automatically collected physical condition data and faces, and children's health status is diagnosed by case-based reasoning by using physical condition data of children. However, only for sick children, the treatment scheme of over-the-counter standing drugs is given by man–machine combination according to their severity. Therefore, on the basis of the proposed training method of intelligent infant health diagnosis and intelligence development, this paper designs an intelligent infant health diagnosis and intelligence development device. This research device has strong reliability and diversified functions, which are beneficial to the infant's health and intelligence development, but there are still some areas that need further research. Even people's personality characteristics, ability tendency and intelligence level are also related to diet. Chinese children's unreasonable diet has various manifestations. According to the literature, the main reason is that the food is too delicate and destroys the content of vitamins and minerals of the food. The relationship between unreasonable dietary structure, nutritional imbalance, and intellectual disability has been recognized by the society, such as iodine deficiency and intellectual disability, and the impact of zinc deficiency on intelligence. However, how to make children with normal development have better intelligence has not attracted extensive attention from the society. The children's perception and cognitive ability, as well as the flexibility and accuracy of children's hands, are trained through three links: self-study of human meridians and acupoints, chapter test, and game breakthrough. Comparative analysis shows that the proposed method can integrate learning, entertainment, and healthcare, and play a role in enhancing the physical and mental health of children and also in developing the children's intelligence.

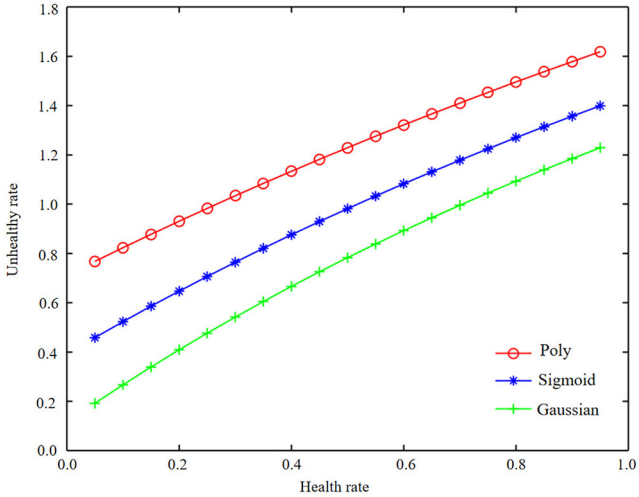


FIGURE 8 | The ROC curve representation of three kernel functions under machine learning combined data.

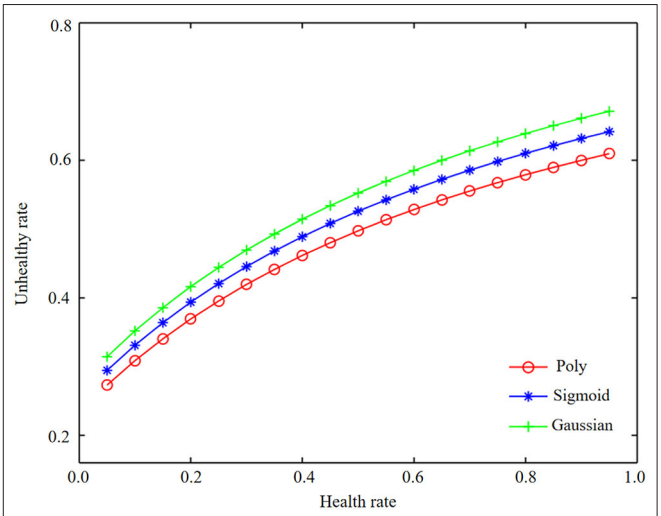


FIGURE 9 | The ROC curve representation of three kernel functions under machine learning combined data.

Results and Analysis

After the original data are processed by SMOTE algorithm, the new dataset is modeled by the same method and its performance is compared. The recall rate, f-score, accuracy, and Area under curve (AUC) of the random forest model are higher than those of the other two models, and the specific results are shown in **Tables 1**, **2**.

Through the investigation of kindergarten a, the specific results of children's physical condition are shown in **Table 3**.

In the process of infant health diagnosis and intelligence development based on machine learning and health information statistics, the accuracy, precision, recalls rate, F1 value, and AUC value are used to evaluate the effects of basic data model and combined data model. By constructing Logistic regression model, the results are shown in **Table 4** and **Figures 3**, **4**.

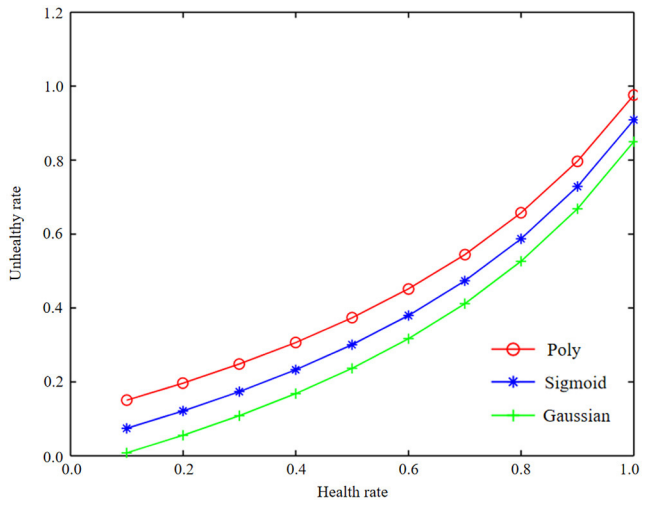


FIGURE 10 | ROC curve performance of three kernel functions under machine learning combined data.

By comparing the modeling results of basic data and combined data, it is found that there is an obvious gap between the modeling results of basic data and combined data. The AUC of the test set of combined data is 83.66%, and that of the test set of basic data is 81.33%. In logistic regression modeling, the AUC of the combined data model is 2.33% higher than that of the basic data, and the accuracy, accuracy, recall F1 values are also greatly improved, indicating that the precision of the model is also improving accordingly. It can be seen from the observation figure that the gap between the AUC results of the training set and the test set is small in the combined data and the basic data, and there is no obvious gap between the accuracy, recall, and F1 value of the training set and the test set of the combined data and the basic data, indicating that there is no overfitting phenomenon in the model.

Using basic data for modeling, the optimal parameters and model evaluation results of different kernel functions are obtained. Get Working characteristic curve (ROC) curves of different kernel function test sets and training sets, as shown in **Figures 5–7**. At the same time, the accuracy, precision, recall rate, F1 value, and AUC value of each model are shown in **Table 5**.

As can be seen from **Figures 5–7**, the test set accuracy of Gaussian kernel function is 74.54%, which is <83.72% of polynomial kernel function, but the recall rate and F1 value of polynomial kernel function are low, indicating that although polynomial kernel function ensures the model precision, the recall rate is low. In addition to the accuracy, other indexes of Gaussian kernel function are the best; especially the AUC of the test set of Gaussian kernel function is 82.57%, which is the largest among all kernel functions. The gap between the training set and the test set of the three SVM kernel functions is small, indicating that there is no fitting phenomenon in the training process. Through comprehensive comparison and analysis, the Gaussian kernel SVM model with C = 1 and gamma = 0.01 is finally selected as the best model for basic data modeling.

Using the combined data to model different kernel functions of SVM, the optimal parameters and model evaluation results of different kernel functions are obtained. The values of accuracy, precision, recall, F1, and AUC are shown in **Table 6**. Meanwhile, ROC curves of different kernel functions are shown in **Figures 8–10**.

As can be seen from **Figures 8–10**, in the machine learning algorithm, the accuracy of Gaussian kernel function is slightly less than polynomial kernel function and Sigmoid kernel function, but other indexes are the best, and the gap between the test set and training set of SVM model of Gaussian kernel function is small, which shows that the model has hardly been fitted. According to the comprehensive comparison of index accuracy, precision, recall rate, F1 value, AUC value, and fitting effect, we finally choose the Gaussian kernel SVM model with C = 1 and gamma = 0.01 as the best model for combined data modeling.

CONCLUSIONS

Children's health level and intellectual development level are the basis of children's other aspects. Although the current products can carry out children's morning examination and daily management of kindergartens, with the development of modern information technology, it is necessary to design a product to integrate children's health examination and intellectual development, as well as learning and entertainment. Based on the training methods of intelligent children's health diagnosis and intelligence development based on machine learning and health information statistics, this paper designs children's intelligent health diagnosis and intelligence development device. The research device of this subject has strong reliability and diversified functions, which is conducive to children's physical health and intelligence development, but there are also places that need to be further studied. Based on machine learning, several basic learners are integrated to train and predict the data, and the prediction effect is usually better than the traditional machine learning model. The method of intelligent infant health diagnosis and man-machine combination for infant disease treatment is proposed. This method recognizes the identity of children through the automatically collected physical condition data and face, uses the child's physical condition data to diagnose the child's health condition based on case-based reasoning, gives the treatment scheme of over-the-counter standing drugs for the sick children according to their severity, and publishes the diagnosis results and treatment suggestions to parents through WeChat platform.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Chengdu University. Written informed consent for

participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SW and ML: conceptualization, methodology, validation, resources, data curation, writing review, and editing. SW: software, formal analysis, investigation, writing original draft preparation, project administration, and funding acquisition. ML: visualization and supervision. All authors have read and agreed to the publishing version of the manuscript.

FUNDING

This work was supported by the Center for Early Childhood Educatoon Research of Sichuan, China (CECER-2020-B03); Social Science Planning Office of Chengdu, China (YN2420210747).

REFERENCES

1. Jones LD, Golan D, Hanna SA, Ramachandran M. Artificial intelligence, machine learning and the evolution of healthcare: a bright future or cause for concern? *Bone Jt Res*. (2018) 7:223–5. doi: 10.1302/2046-3758.73.BJR-2017-0147.R1
2. Li C, Rusak Z, Horvath I, Ji L. Development of engagement evaluation method and learning mechanism in an engagement enhancing rehabilitation system. *Eng Appl Artif Intell*. (2016) 51:182–90. doi: 10.1016/j.engappai.2016.01.021
3. Yan K, Jin Q. Guest Editorial: AI and Machine Learning Solution Cyber Intelligence Technologies: New Methodologies and Applications. *IEEE Trans Ind Inform*. (2020) 16:6626–31. doi: 10.1109/TII.2020.2988944
4. Andrysiak T. Machine learning techniques applied to data analysis and anomaly detection in ECG signals. *Appl Artif Intell*. (2016) 30:610–34. doi: 10.1080/08839514.2016.1193720
5. Weiner MG. POINT: Is International Statistical Classification of Diseases and Related Health Problems, 10th Revision Diagnosis Coding Important in the Era of Big Data? Yes. *Chest*. (2018) 153:1093–95. doi: 10.1016/j.chest.2018.01.025
6. Ismail L, Materwala H, Tayefi M, et al. Type 2 Diabetes with artificial intelligence machine learning: methods and evaluation. *Arch Comput Methods Eng*. (2021) 26:1—21. doi: 10.1007/s11831-021-09582-x
7. Liu Q, Liu Q. Prediction of volleyball competition using machine learning and edge intelligence. *Mobile Inform Syst*. (2021) 2021:1–8. doi: 10.1155/2021/5595833
8. Narudin FA, Feizollah A, Anuar NB, Gani A. Evaluation of machine learning classifiers for mobile malware detection. *Soft Comput*. (2016) 20:343–57. doi: 10.1007/s00500-014-1511-6
9. Wang J, Zhang L, Duan L, Gao RX. A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *J Intell Manuf*. (2017) 28:1125–37. doi: 10.1007/s10845-015-1066-0
10. Prashanth R, Roy SD, Mandal PK, Ghosh S. High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. *Int J Med Inform*. (2016) 90:13–21. doi: 10.1016/j.ijmedinf.2016.03.001
11. Sepúlveda NFE, Sinha JK. Blind Application of Developed Smart Vibration-Based Machine Learning (SVML) Model for Machine Faults Diagnosis to Different Machine Conditions. J Vib Eng Technol. (2020) 13:1–10.
12. Nina S, Brian W. Artificial intelligence and the future of global health. *Lancet*. (2020) 395:1579–86. doi: 10.1016/S0140-6736(20)30226-9
13. Rahman MM, Khatun F, Uzzaman A, Sami SI, Bhuiyan MA, Kiong TS. A comprehensive study of artificial intelligence and machine learning approaches in confronting the coronavirus (COVID-19) pandemic. *Int J Health Serv*. (2021) 2021:1–16. doi: 10.1177/00207314211017469
14. Buchlak QD, Esmaili N, Leveque JC, Bennett C, Farrokhi F, Piccardi M. Machine learning applications to neuroimaging for glioma detection and classification: an artificial intelligence augmented systematic review. *J Clin Neurosci*. (2021) 89:177–98. doi: 10.1016/j.jocn.2021.04.043
15. Tai AMY, Albuquerque A, Carmona NE, Subramanieapillai M, Cha DS, Sheko M, et al. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artif Intell Med*. (2019) 99:101704.1°101704.11. doi: 10.1016/j.artmed.2019.101704
16. Vellido A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput Appl*. (2019) 39:1–15.
17. Yao H, Mai T, Jiang C, Kuang L, Guo S. AI Routers & Network Mind: A Hybrid Machine Learning Paradigm for Packet Routing. *IEEE Comput Intell Mag*. (2019) 14:21–30. doi: 10.1109/MCI.2019.2937609
18. van der Schaar M, Alaa AM, Floto A, Gimson A, Scholtes S, Wood A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Mach Learn*. (2021) 110:1–14. doi: 10.1007/s10994-020-05928-x
19. Kodesh A, Levine SZ, Khachadourian V, Rahman R, Schlessinger A, O'Reilly PF, et al. Maternal health around pregnancy and autism risk: a diagnosis-wide, population-based study. *Psychol Med*. (2021) 108:1–9. doi: 10.1017/S0033291721001021
20. Fernandez Escamez CS, Martin Giral E, Perucho Martinez S, Toledano Fernandez N. High interpretable machine learning classifier for early glaucoma diagnosis. *Int J Ophthalmol*. (2021) 14:393–8. doi: 10.18240/ijo.2021.03.10
21. Johnson NT, Andi D, Hughes KJ, Korkin D. Biological classification with RNA-seq data: Can alternatively spliced transcript expression enhance machine learning classifiers? *RNA*. (2018) 24:1119–32. doi: 10.1261/rna.062802.117
22. Olusanya BO, Wirz SL, Luxon LM. Community-based infant hearing screening for early detection of permanent hearing loss in Lagos, Nigeria: a cross-sectional study. *Bull World Health Organ*. (2008) 86:956–63. doi: 10.2471/BLT.07.050005
23. Yousef N, Singh Y, Luca DD. "Playing it SAFE in the NICU" SAFE-R: a targeted diagnostic ultrasound protocol for the suddenly decompensating infant in the NICU. *Eur J Pediatr*. (2021) 64:1–6. doi: 10.1007/s00431-021-04186-w

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Li and Ng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fpubh.2022.915801

Analysis on the Cognitive Impact of Social Mobile Games on Left-Behind Children in the Era of Big Data

Mengyu Li[1,2], *Jinglei Li*[1]*, *Megat Al Imran Yasin*[2], *Mohd Nizam Osman*[2], *Norliana Binti Hashim*[2], *Lay Hoon Ang*[2] and *Yitian Xue*[3]

[1] School of Journalism and Communication, Zhengzhou University, Zhengzhou, China, [2] Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Malaysia, [3] Propaganda Department, Hohai University, Nanjing, China

The popularity of mobile gaming has become a common sight in rural areas, and the problem of left-behind children's mobile gaming has become the biggest challenge faced by rural education, and has become a hot topic in the education sector and society. The stage of left-behind children is the golden period of learning and growth. However, this stage is also the period when they have the greatest fluctuations in their ideology due to various factors such as guardians and youths. With the development of big data, it has been applied to various aspects of people's life. This article is mainly based on qualitative research, with the interview and observation as the main methods, supplemented by a questionnaire survey method. In the empirical analysis part, this article has a certain degree of expansion in data selection and research methods. Compared with previous studies, we increased the scale of the research data, making the research results more meaningful. In the research method, the ordinary least squares method (OLS), the propensity score matching method (PSM) and the two-stage least square method (2SLS) are used, and multiple control variables are selected. The factor analysis of the original scores of the historical knowledge test and the original scores of the two-dimensional cultural value evaluation are carried out to obtain the factor scores of cognitive ability. The emotion of the child is expressed by the depression score, and factor analysis is also performed on the depression score. Cognitive abilities refer to reasoning or thinking, processing speeds, and one's ability to solve problems in novel situations, independent of acquired knowledge. OLS regression results show that left-behind children are inferior to non-left-behind ones in cognitive ability. Moreover, left-behind children are more likely to be emotionally depressed. And whether the children are accompanied by migrant or rural parents, there is no significant difference in their cognitive ability and emotions. Because there are unobservable factors that affect whether children are left-behind and children's cognitive abilities and emotions, the sample may have a self-selection bias. This research focuses on the phenomenon of left-behind children's mobile gaming, revealing the compromised cognitive abilities of these marginalized children groups. Our study might put a wake-up for authorities on the education in rural areas.

Keywords: rural economy, social mobile games, left-behind children, cognitive impact, big data

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Srinivasa Rao Namburi,
Bapatla Engineering College, India
Praveen Kumar,
VIT University, India

***Correspondence:**
Jinglei Li
ljinglei@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *08 April 2022*
Accepted: *02 May 2022*
Published: *06 June 2022*

Citation:
Li M, Li J, Yasin MAI, Osman MN, Hashim NB, Ang LH and Xue Y (2022) Analysis on the Cognitive Impact of Social Mobile Games on Left-Behind Children in the Era of Big Data. Front. Public Health 10:915801. doi: 10.3389/fpubh.2022.915801

INTRODUCTION

In recent years, various mobile games have been surging and swept across the country with strong momentum. As of 2019, the number of mobile Internet users in China was 847 million, an increase of 29.84 million from the end of 2018. The proportion of Internet users using mobile phones to access the Internet has increased by 99.1% from 98.6% at the end of 2018 (1). Among the netizens, the students are the most. Among them, the netizens with junior high school, high school/technical school/technical school degrees accounted for 38.1 and 23.8%, respectively; the netizens with college, university, and undergraduate education accounted for 10.5%, respectively. It can be seen that junior high school students use the Internet more often (2). As of June 2019, the number of national online game users reached 494 million, an increase of 9.72 million from the end of 2018, accounting for 57.8% of the total netizens; the number of mobile online game users reached 468 million, an increase of 8.77 million from the end of 2018, accounting for 55.2% of mobile netizens (3). From this point of view, mobile Internet users will play some mobile games to a certain extent, and mobile games are very popular among junior high school students. The survival of marginalized children's mobile gaming has become a common sight in rural areas, and the problem of marginalized children's mobile gaming has become the biggest challenge faced by rural education, and has become a hot topic in the education sector and society (4). Marginalized children incorporate hand-held games into their lives. Hand-held games have become a way of life and another living space for them. In the process of playing hand-held games, they have generated their own culture and civilization, forming a distinctive two-dimensional cultural landscape (5). However, the outside world's understanding of mobile games is still living under moral "panic." Therefore, mobile games have raised various criticisms, ignoring their positive significance and inherent cultural value of mobile games. This research cuts into the phenomenon of mobile games of left-behind children in rural areas from the perspective of the two-dimensional culture, which has theoretical penetrating power and profoundness (6).

The mobile game survival phenomenon of left-behind children in rural areas is a product of the times and a generation that has been "sacrificed" under social structural reforms. At the same time, the phenomenon of marginalized children's mobile game survival is the primary issue that rural education needs to face directly, and requires sufficient attention from the education sector (7). For mobile games, we should get rid of all kinds of prejudices and panic, look at new things with a rational and dialectical perspective, attach importance to their derived cultural functions historical values, understand their inner spiritual needs and respect their cultural choices and the right to play, and what needs to be done is to guide rather than prohibit. The evolution of civilized and restrained hand demonstrations of left-behind children in rural areas requires a multi-faceted effort. Family education and school education are the top priorities, and then changes in the social structure, social environment, and ethos are the most important to solve the root of the problem (8). At the same time, there are data showing that the playtime of left-behind children is significantly higher than that of non-left-behind children: "playing 4–5 hours a day" accounted for 18.8 and 8.8%, respectively, and "playing more than 6 hours a day" accounted for 18.8% respectively (9). Therefore, marginalized children have become the main force among young players. The survival of marginalized children in mobile games has aroused the attention of all parties and the appeal of experts, setting off a wave of "turbulent waves." Under the panic of "playing mobile games is sapping the spirit" the government, experts, society, and distressed parents tried to find various measures, but with little effect (10).

The mobile game survival phenomenon of left-behind children in rural areas has attracted attention from all quarters. However, by searching the literature on mobile games and the two-dimensional culture, it was found that the two overlapped very little. The factor analysis of the historical knowledge level and the two-dimensional cultural value evaluation are carried out to obtain the factor scores of cognitive ability. The emotion of the child is expressed by the depression score, and factor analysis is also performed on the depression score. Therefore, this research analyzes the behavior of the marginalized children's hand march from sociology, and takes an insight into the fetters of the marginalized children and mobile games from the perspective of the two-dimensional culture, explores the deep-level educational sociology of the phenomenon, and at the same time reveals the mobile games. The nature of the phenomenon and the cultural connotation it contains not only provide theoretical results for the research on the phenomenon of left-behind children's mobile games, but also inject blood into the theory of the two-dimensional culture and keep pace with the times. Through the investigation and research on the phenomenon of left-behind children's mobile games, it helps us understand the deep-seated needs and living conditions of left-behind children's mobile game players in the virtual game zone, to better guide the gaming behaviors of left-behind children, and at the same time, to provide a new way out and a new reference for solving the problem of left-behind children's mobile game addiction, and to trigger thinking about rural education and the education of left-behind children.

RELATED WORK

Some scholars' research on the two-dimensional culture and film is basically focused on aesthetics. They believe that two-dimensional culture changes the aesthetic style of domestic films. This change is manifested in: first, the characterization of characters. The characters are set with "cute" as the element and tend to be soft and beautiful. Second, time and space reconstruction. Using virtual reality (VR) technology to create a virtual space-time, this space-time can travel through different eras, and achieve the ideal utopia for human beings. Third, the method of film narration uses ACG elements. Of course, there are also changes in audiovisual language, values, and plot settings. Wang (11) believes that the two-dimensional culture has an impact on the concept, audience change,

and aesthetic aspects of film and television creation. Recent advances in information technologies and data science have enabled convenient access, storage, and analysis of massive on-site measurements, bringing about a new big-data-driven research paradigm (22). With the development of Internet of Things (IoT), 5G, and cloud computing technologies, the amount of data from manufacturing systems has been increasing rapidly (23).

In terms of film aesthetics, Fu and Zhao (12) elaborated on the influence of the two-dimensional culture on the aesthetic tendency of films from three aspects: character image, time and space landscape, and value concept. Next, Li et al. (13) believed that the two-dimensional culture has the effect of reducing the aesthetic cracks in the film aesthetics and enhancing the value of the characters, but at the same time it also has the negative effect of excessive commercial consumption and disintegrating the character archetypes. He believes that the two-dimensional culture has brought unique creation to the film. The method also brings a relaxed, cheerful, vigorous and youthful values that are different from the traditional ones, and establishes a unique aesthetic outlook. Niu et al. (14) inherited the research of predecessors and elaborated on film techniques, film styles and film values, and proposed a kind of aesthetic and healing film aesthetic values. He believed that under the influence of the two-dimensional culture, the narrative method and aesthetics of the film have changed a lot, and the two-dimensional culture is the source of motivation for creating excellent movies to integrate the two and avoid being symbolized and conceptualized as the research focus. Researchers believe that the two-dimensional culture has a negative impact on college students who are self-proclaimed and detached from society. On the basis of questionnaire surveys, they understand their psychological needs, and put forward corresponding countermeasures and opinions from schools, families and themselves (18). Li et al. (15) explained that the "cute" factor in the secondary culture can dispel the "scars" and "cruel" gray tones of traditional youth movies, blend the aesthetics of business, culture, and audiences, and then propose movie narratives which is influenced by the two-dimensional culture. It has a personality and reflects the unique aesthetic style of youth. At the same time, in the pursuit of youth, there is a mentality of escaping from reality and confusion, and a cultural mentality that is both lost and recognized. Regarding the two-dimensional culture to promote the development of movies, Liu and Zhou (16) proposed that according to the characteristics of the two-dimensional culture, it should pay attention to learning from and absorbing its outstanding characteristics, and on this basis, form its own unique style to promote the creation and development of film and television.

In terms of countermeasures to reduce the impact of the two-dimensional culture on students, Yue et al. (17) believes that the two-dimensional virtual culture has unique characteristics, attracting a wide range of young people to join, and also leading to the emergence of primary school students' "Internet addiction": improving laws and strengthen supervision; parents strengthen education, supervision and care for their children; schools strengthen quality education and enrich after-school

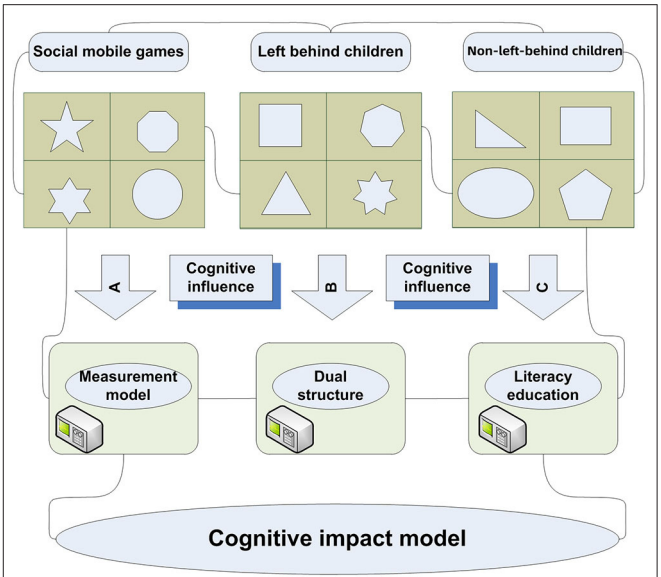


FIGURE 1 | Schematic diagram of the cognitive measurement model of left-behind children.

life; providing a good growth environment. The research is more specific-self-awareness. It is proposed that the different types of secondary culture and the length of contact time have different effects on the development of self-awareness of junior high school students. The corresponding countermeasures are presented: updating concepts and improving media literacy awareness; secondary media self-discipline, ingenious two-dimensional resources; forming educational synergy to promote media literacy education; giving full play to the main role of junior high school students to improve students' self-awareness level (19, 20). In terms of information dissemination, some scholars have their own unique opinions. They believe that under the influence of the two-dimensional culture, the information they choose is to blur the boundaries of countries, ethnicities and regions, but at the same time they have a strong patriotic sentiment and pay attention to national current affairs. However, they remain skeptical of the official information, and hope that the Internet will answer this. Moreover, they remain calm about the information and do not blindly follow the trend, but they pay more attention to their own interests (21).

CONSTRUCTION OF A MODEL OF THE COGNITIVE IMPACT OF SOCIAL MOBILE GAMES ON LEFT-BEHIND CHILDREN

Cognitive Levels of Left-Behind Children

After several developments, the original connotation is no longer there, but the group of left-behind children has received more and more attention. Left-behind children are the product of the urban-rural dual structure. They are both disadvantaged groups and "marginal people." Their educational problems and

generation methods touch the nerves of society and are difficult to ignore. **Figure 1** shows a schematic diagram of the cognitive measurement model for left-behind children.

The first measurement model used in this article is ordinary least square (OLS). Ordinary least squares estimation is to find the estimated values of the parameters, so that the sum of squared deviations of the below formula can be minimized. In the formula, the weight of each square term is the same, which is the parameter estimation method of ordinary least squares regression.

$$X = \{x(t) \in R | t = 1, 2, 3, .., T\} \tag{1}$$

The explanatory variable is a definite variable, not a random variable; the random error term has zero mean, same variance, the random error term is not correlated with the explanatory variable; the random error term obeys zero mean, same variance, and zero covariance normally distributed, ordinary least squares estimation is a linear unbiased estimation of the minimum variance of the regression parameters.

$$p(X|t) = \sum_{t} p(X|z, t)p(t|x) \tag{2}$$

Among them, y is the explained variable, including children's cognition and depression (emotion), children is the attention variable, which represents the type of child. Taking the non-left behind children as the benchmark group, we construct left-behind children relative to non-left behind and migrant children relative to each other. As a binaryz duzmmy variable for non-left behind children, X is the control variable, including the child's gender, child's age, education investment in the child in the past year, mother's childbearing age, father's childbearing age, mother's years of education, and father's years of education and per capita net income of the household last year, it is a random interference term.

$$p(C|X) = \prod_{s}^{T} p(c(s)|c(1), c(2), ..., c(s-1), X) \tag{3}$$

In order to control the self-selection problem, this paper introduces a propensity score matching model. The basic idea is to find similar individuals in the treatment group (left-behind children/migrated children) in the control group (non-left behind children). The probability of these individuals becoming left-behind children/migrated children is similar and comparable, and can be matched and estimated.

$$t = \ln sig(\overline{s(x)} - s(\bar{x})) = \frac{1}{1 + \exp(\overline{s(x)} - s(\bar{x}))} \tag{4}$$

Specifically, the binary treatment variable divides the children in the sample into two groups. For the child individual i ($i = 1, 2, \ldots, N$), it means that the child is a non-left behind child/migrated child (treatment group), it means that the child is a non-left behind child (control group). There are two potential outcome variables, $x(0)$ and $x(1)$. As shown in formula, the average processing effect of left-behind children/migrating children interested in this article on cognitive level and emotion is:

TABLE 1 | Cognitive level distribution table of left-behind children.

Group index	Historical knowledge	Two-dimensional cultural value evaluation	Overall Cognitive ability score	Depressive emotion score
1	0.261	0.433	0.058	0.228
2	0.441	0.667	0.095	0.366
3	0.221	0.352	0.048	0.178
4	0.132	0.423	0.065	0.281

$$x(t) = \begin{cases} t, 0 < t < 1 \\ a * s(t) + (1 - a) * s(t - 1), t > 1 \end{cases} \tag{5}$$

Among them, $E[Y(1)|D(i) = 1]$ is the counterfactual mean value that cannot be obtained from the observable information. Generally, $E[Y(1)|D(i) = 0]$ can be used instead, but this method will cause self-selection bias due to factors that affect the type of children and also affect cognitive ability and emotion. The idea of propensity matching score can solve this selective bias.

$$y(x) - (a(0) + a(1) * c(t) + a(2) * x(t)) = 0 \tag{6}$$

The propensity matching score needs to meet two conditions. The first hypothesis is the conditional independence hypothesis. The content of this hypothesis is that given a series of control variables, the value of the result variable $Y(i)$ is independent of the binary processing variable $D(i)$. The second hypothesis is the common support, that is, the value ranges of individual propensity scores of children with the characteristics of the control variables overlap. If the assumptions are meet, such as the formula, the average processing effect of left-behind children/migrated children on cognitive level and emotion is:

$$A(X) = E[Y(1)|D(i) = 1] - E[Y(0)|D(i) = 1] \tag{7}$$

Since whether a child is a left-behind child or a migrant child is affected by the choice of parents who go out to work, at the same time, whether the parents go out to work will also affect the child's cognition and emotion, the ordinary least squares method has endogenousness. **Table 1** shows the cognitive level distribution table of left-behind children. Therefore, this paper introduces the child type instrumental variable—the proportion of the labor force in the village who go out to work in the total labor force, and uses the two-stage least squares method (2SLS) to solve the endogenous problem of the model.

TABLE 2 | Analysis of operating elements of social mobile games.

Variable item	Constant term	Sample size	*R*-squared
A	1.793	2,355	0.343
B	2.384	2,411	0.289
C	3.012	2,561	0.336
D	2.774	2,713	0.257
E	3.061	2,615	0.281

Operating Elements of Social Mobile Games

A narrowly defined mobile game refers to a form of the game that can only be run on mobile phones. In a broad sense, mobile games refer to a form of games that can be run on mobile phones and various devices. With mobile phones as the carrier, it is mobile, so it is also called mobile games. With the popularity of mobile phones and the Internet, mobile games have become more popular than ever. There are many types of mobile games and various fancy styles, but Multiplayer Online Battle Arena (MOBA) mobile games are strong and popular among gamers. Among them, "Honor of Kings" is a milestone in MOBA mobile games, pushing MOBA mobile games to a climax. MOBA, also known as a tactical competitive game, is based on strategy and teamwork, emphasizing competition and cooperation. "Honor of Kings" combines role-playing, battlefield kills, intellectual victory, teamwork, and social attributes into one. It not only meets the needs of players for "happy enmity" and a vivid battlefield, but also meets friends and emotions of the need for communication. **Table 2** shows the analysis of operating elements of social mobile games.

In people's communication activities, meaningful gestures have begun to work. Significant gestures related to the use of symbols always presuppose that everyone involved in communication activities has the ability to imagine their actions from the standpoint of others and to play the role of another. In symbol communication, people interpret each other's attitudes and actions based on the meaning produced by the interpretation of symbols. The process of human communication involves actors constantly making self-conscious adjustments to the actions of others, that is, repeatedly adapting to each other's actions through definition and redefinition, interpretation and reinterpretation. The application of symbolic interaction theory in this research is embodied in two aspects. On the one hand, the interaction between left-behind children in rural areas and groups in the life world is affected by symbolic interaction theory, including both peer groups and their guardians; on the other hand, the interaction between children and the group in the network virtual space is influenced by the theory of symbolic interaction, and the group mainly includes managers and other game players. When a teammate shows contempt for the equipment of left-behind children in the countryside, they tend to spend more money to buy the equipment. When other players pursue a certain hero, the marginalized children will make the same approach as other game players, and tend to choose the hero.

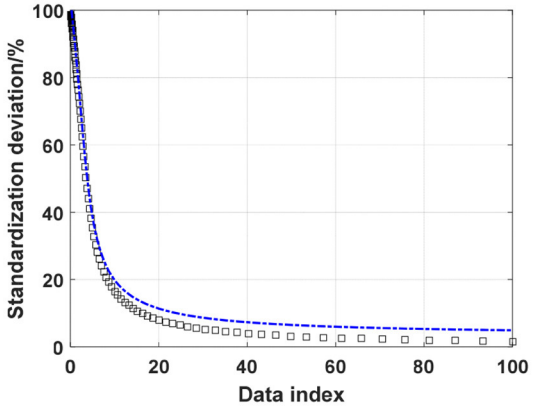


FIGURE 2 | The distribution of the standardized deviation curve of the cognitive impact model.

Cognitive Impact Model Optimization

The mobile game market has become popular recently, which has brought huge economic benefits and also caused a certain negative impact. Among them, marginalized children, as a vulnerable group, are deeply affected when they come into contact with this mobile game. Therefore, this article first defines related concepts such as "left-behind children in rural areas" and "cognition." The research object is identified as left-behind children in rural areas. Through in-depth interviews with them, we can understand their views on mobile games and their attitudes toward the historical figures in the design. In order to dig out the impact of mobile games on the cognition of left-behind children in rural areas, we further analyze the reasons for this impact, and propose practical and feasible measures to resolve the negative impact. Secondly, we carry out relevant literature collection and draw on the research of related scholars on games and marginalized children. In the model estimation of the left-behind children, the standardized deviations after the control variables were matched were all reduced to $<10\%$ except for the quadratic term of the father's age, and the t-test results were not significant, that is, the model estimation of the left-behind children satisfies propensity matching to score the assumption that the control variables of the treatment group and the control group are balanced. **Figure 2** shows the distribution of the standardized deviation curve of the cognitive impact model. In the model estimation of left-behind children, the standard deviation of most control variables after matching fell below 10%. Although the per capita net income of the family is above 10%, it has dropped a lot from the 24.3% before matching. We satisfy the assumption that the control variables of the propensity matching score treatment group and the control group are balanced.

This article mainly adopts qualitative research methods, including the interview method, and observation method. We read a large number of literature materials, based on previous

studies, through induction, comparison, analysis and other steps, to understand the current academic research on the impact of mobile games on marginalized children's cognition abilities, and put forward our research direction, and refer to relevant research results in a targeted manner. It is a face-to-face purposeful conversation between the researcher and the researchee. According to whether the interview is structured or not, it is divided into structured interviews, semi-structured interviews and unstructured interviews. This study mainly adopted a semi-structured interview method, and mainly selected six marginalized children as interview subjects. During the survey process, as participants, we carefully observe the emotional expression of marginalized children during interviews, and observe the family environment of the left-behind children in rural areas to understand the family's educational atmosphere, and observe their verbal expressions when answering questions to prove the authenticity of the answers.

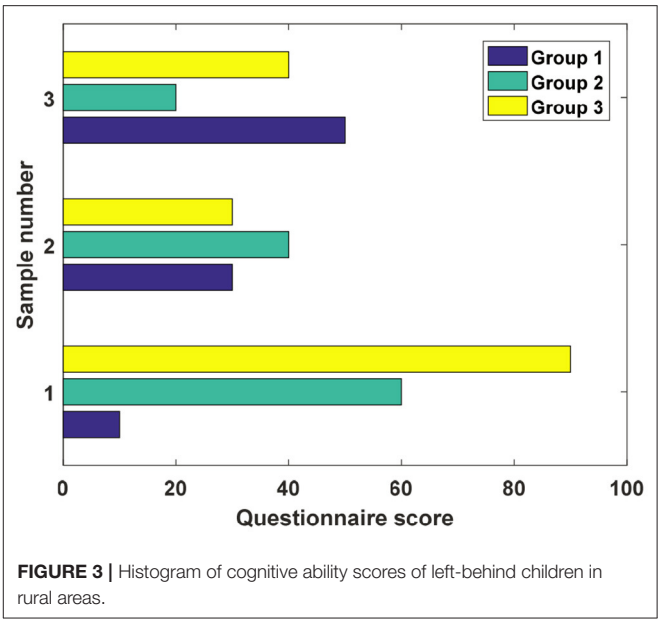


FIGURE 3 | Histogram of cognitive ability scores of left-behind children in rural areas.

APPLICATION AND ANALYSIS OF THE MODEL OF SOCIAL MOBILE GAMES' COGNITIVE IMPACT ON LEFT-BEHIND CHILDREN

Cognitive Analysis of Left-Behind Children in Rural Areas

According to the research objectives, this article mainly focuses on the development of questionnaires for left-behind children in rural areas. Distributed 466 questionnaires for students, 349 for left-behind children, 117 for non-left-behind children, excluding 8 doodles and incomplete questionnaires, and 115 valid questionnaires for non-left behind children were recovered. The recovery rate was 98.2%. There were 343 left-behind children, and the recovery rate was 98.3%. It also conducts a scientific and systematic analysis of the statistical data, provides a factual basis for the education status of left-behind children in rural areas in the region, and provides microdata support for the study. This article mainly compiles the questionnaire from the basic structure of the family of left-behind children and non-left behind children, the strength of family nurturing functions, and the children's performance in historical knowledge level and two-dimentional cultural value evaluation. The basic structure of the child's family includes who the guardian is at home, the education level of the guardian. Whether it is an only child, the parenting function of the family includes the care and supervision of guardians and parents of migrant workers in their children's historical knowledge level and two-dimensional cultural value evaluation, which includes the level of knowledge about historical figures and stories (such as narrative, ancient poetry, extracurricular reading, etc.), as well as attitudes and evaluations related to two-dimensional culture. **Figure 3** shows the histogram of the cognitive assessment scores of left-behind children in rural areas.

It can be seen that about 86.1% of left-behind children are taken care of by their grandparents (grandparents or grandparents), about 8.7% of left-behind children are taken care of by uncles and uncles (or relatives), and left-behind children living alone account for 5.2% of the total number of left-behind children. After controlling for the endogenity of the model, the type of child has no substantial effect the performance in historical knowledge level. The significant effect in the OLS model is due to the inclusion of other uncontrolled unobservable factors. The gender of children has no significant effect on their original scores on the performance in historical knowledge level, but for every year the children's age increases, their original scores on historical knowledge test increase by 1.447 on average, and this result is significant at the 1% level. **Table 3** shows the standardized deviation distribution of the control variables before and after model matching.

There is no significant impact on children's educational investment, mother's childbearing age, mother's childbearing age square item, father's childbearing age square item, mother's years of education, and children's performance in historical knowledge level test raw scores. The original score on the historical knowledge test was reduced by 0.275, and this result was significant at the 10% level. For every additional year of the father's education years, the original score on the children's historical knowledge level test increases by an average of 0.195, and this result is significant at the 1% significance level. Household net income per capita has no significant effect on the raw scores of children's historical knowledge test.

Evaluation of Social Mobile Games in the Rural Economy

Because of regional differences and informants are affected by some objective conditions, the subjects of the interview program are only students in 11 classes from the second to sixth grade of elementary school, of which there are two classes in each of the second, third, fourth, and fifth grades, 51 informants were obtained by purposeful sampling. The number of each class are:

TABLE 3 | Standardized deviation distribution of control variables before and after model matching.

Attributes number	Group A		Group B	
	Before matching	After matching	Before matching	After matching
Age	9.5	5.8	2.4	1.7
Gender	4.8	7.2	5.9	6.7
Educational investment	12.2	11.2	2.9	3.6
Net income per capita	5.6	7.4	11.6	9.2

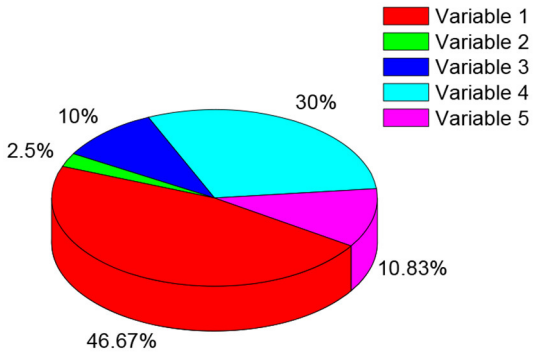


FIGURE 4 | Cognitive abilities distribution of left-behind children among the explained variables.

44 in class 2 (1) and 11 left-behind children; 47 in class 2 (2) and 12 left-behind children; 41 in class 3 (1) and 8 left-behind children; class 4 (1) with 46 people, 10 left-behind children; class 4 (2) with 45 people, 14 left-behind children. There are 42 students in the class and 12 left-behind children. There are three classes in grade six, each with 40 students in class six (1) and nine left-behind children; 39 students in class six (2) and 10 left-behind children; 40 students in class six (3) and 13 left-behind children.

Figure 4 shows the cognitive fan distribution of left-behind children among the explained variables. The child's cognition in the explained variable uses a set of two-dimensional culture value questions and a set of historical questions to test and evaluate the cognitive level of all respondents who need to complete the self-answer questionnaire (that is, children aged from 10–15 years old). The factor analysis of the original scores of the historical test and the original scores of the two-dimensional cultural evaluation test is carried out to obtain the factor scores of cognitive ability. This article's measurement of children's emotions (depression) combines six related questions in the questionnaire: "How often did you feel depressed, and unable to do anything when you played "Honor of Kings"? How often did you feel nervous when you played "Honor of Kings"? How often did you feel restless and difficult to stay calm when you played "Honor of Kings"? How often did you feel hopeless in the future when you played "Honor of Kings"? How often did you find it difficult to do anything when you played "Honor of Kings"? When you played "Honor of Kings", do you think life is meaningless?" According to their own circumstances, the interviewees chose the answer that best suits them among the answers: "never, sometimes, half the time, often, almost every time." For answering option: "never = 1 point; sometimes = 2 points; half the time = 3 points; often = 4 points; almost every time = 5 points". This article uses two measurement methods. The first is to add the answers to the six questions to the depression score, and the second is to perform factor analysis on the answers to the six questions to obtain the factor score of depression.

Example Application and Analysis

It can be seen that the average original score of the historical knowledge test for the full sample is 10.51 points, that is, 10.51 of the 24 questions for non-left children with registered permanent residence are answered correctly, but the standard deviation is 4.53, and the fluctuation is large. A total of 34 test questions were designed for the two-dimensional cultural value evaluation. On average, 20.81 questions were answered correctly by non-left children, but the standard deviation was 7.51, and the scores between children were quite different. The children's total scores for the 6 emotional questions answered were 26.99 points. The overall children interviewed had a more positive emotional state. **Figure 5** shows a line graph of cognitive ability factor and scores depression factor scores for different data points. Cognitive ability factor scores and depression factor scores are derived from factor analysis, and the mean value is 0, and the variance is both 1.

Among the interviewed children, non-left children, left-behind children, and migrant children accounted for 31, 49, and 20%, respectively. The proportion of non-left children was very high, and the proportion of left-behind children was the lowest. Among the interviewed children, boys accounted for 53% and girls accounted for 47%. There were slightly more boys than girls, all aged from 10–15 years old. The average education investment per child is 1,600 yuan per year, and the average education investment is at a relatively low level, and the standard deviation is 2.65, which is quite fluctuating. The average childbearing age of the interviewed mother is 26.34 years old, and the father is 28.03 years old. The average childbearing age of fathers is slightly higher than that of mothers. The average childbearing age is within the normal range, but the minimum childbearing age of the interviewed mothers is 20 years old and the youngest father is 22 years old. Similarly, some villagers do not have a scientific

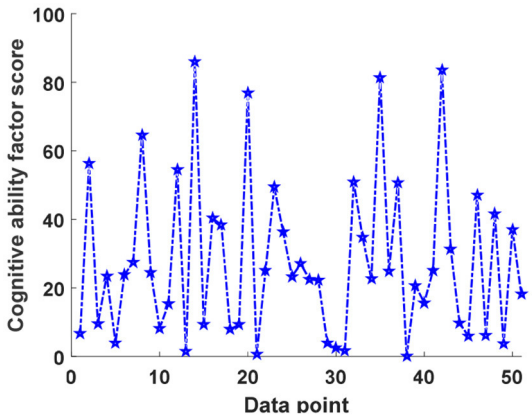


FIGURE 5 | Line chart of cognitive ability factor scores and depression factor scores for different data points. The *symbol indicates the child who takes the cognitive ability test. Y-axis indicates score. There are 51 children in the test.

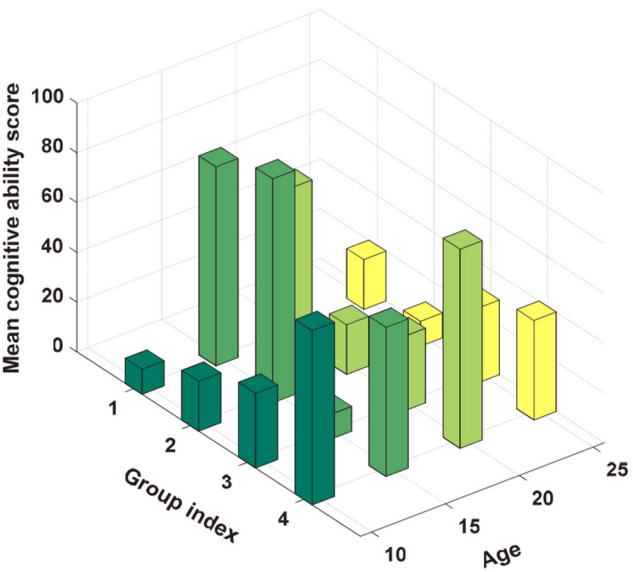


FIGURE 6 | A three-dimensional histogram of the average cognitive ability scores and depression factor scores of left-behind children in different age groups.

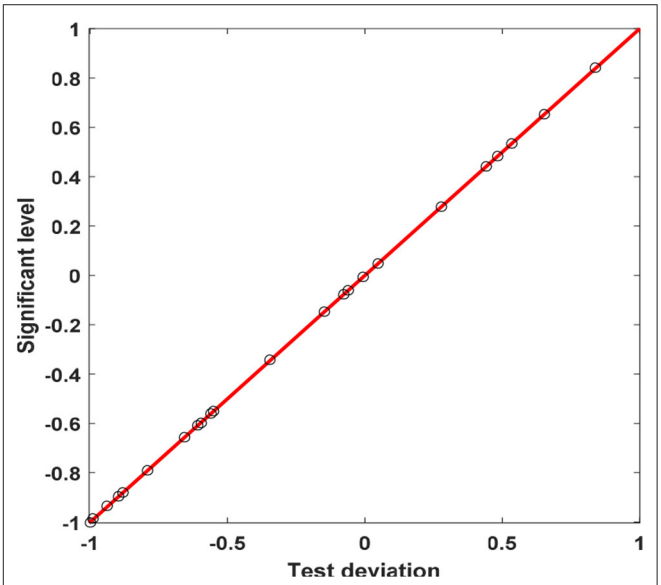


FIGURE 7 | Significance level distribution of cognitive ability test of left-behind children.

understanding of the issue of childbearing age. The mother's average education years is 8.22 years, and the father's 9.82 years. The parents' educational level is relatively low, and the grandparents' education years are at least 1 year, which is basically equivalent to ignorance. This will definitely have a negative impact on the children's cognition. Parents with low educational level generally get married and have children earlier. The average per capita net income of households is 6,350 yuan, which is much lower than the urban per capita income level, indicating that the national urban-rural income gap is still not optimistic. The standard deviation is 5.97, indicating that the per capita net income of the interviewed households is also very different.

The average proportion of the labor force working outside the village is 33.48%. There is a large gap in the proportion of the labor force going out in different regions. This is directly related to the imbalance in national economic development. It is also mentioned in the analysis of the current situation of children that the economically backward central and western regions have a relatively high proportion of migrant workers, while the economically developed eastern regions are much lower. Left-behind children are also mainly concentrated in the central and western regions, and in the economically developed regions. **Figure 6** shows a three-dimensional histogram of the average cognitive ability scores of left-behind children in different age groups. The average depression scores of non-left behind children, left-behind children, and migrant children were 77.11, 86.37, and 66.96, respectively. From descriptive statistics, it can be seen that non-left behind children rarely have depression, followed by children with migration. Left-behind children are most likely to have depression emotions. The depression score is the direct sum of the interviewed children's scores for answering six questions, and the depression factor score is a factor analysis, which is equivalent to standardizing the interviewed children's depression scores, which can be seen from descriptive statistics. The scores of depression factors of left behind children are the highest, and non-left behind children are the lowest.

Figure 7 shows the distribution of the significance level of the cognitive ability test for left-behind children. On the premise that other conditions remain unchanged, the average original score of left-behind children in the two-dimensional cultural value test is 0.585 lower than that of non-left children, but this result is not statistically significant. On average, the original scores

of male children's two-dimensional cultural value evaluation tests are 1.308 lower than female children, and this result is significant at the 5% level. However, gender has no significant effect on the original scores of two-dimensional cultural value test. Every time the child's age increases by 1 year, the original score of the two-dimensional cultural value test will increase by 1.435 on average, and this result is significant at the 1% significance level. For every additional 1,000 yuan invested in children's education, their original two-dimensional cultural evaluation scores increase by 0.195 on average, and this result is significant at the 5% level of significance. For every additional year of mother's years of education, the original score of the children's two-dimensional cultural value test increases by an average of 0.177, and this result is significant at the 5% level. For every year of father's education, the average increase of the original score of the children's two-dimensional cultural value test is 0.458, and this result is significant at the 1% significance level, because fathers generally pay more attention to children's learning. For every 1,000 yuan increase in family net income per capita, the original score of children's two-dimensional cultural value test increases by 0.032 on average, but this result is not statistically significant.

CONCLUSION

The stage of left-behind children is the golden period of learning and growth. However, this stage is also the period when they have the greatest fluctuations in their ideology due to various factors such as guardians and youths. The popularity of mobile gaming has become a common sight in rural areas, and the problem of marginalized children's mobile gaming has become the biggest challenge faced by rural education, and has become a hot topic in the education sector and society. This research focuses on the impact of mobile games on the cognition of left-behind children in rural areas, the reasons for the impact, and suggestions on how to solve the negative impact. Through research, it is found that mobile games have two main influences on the cognition of left-behind children in rural areas: first, the positive impact is conducive to expanding the knowledge reserves of historical figures and two-dimensional culture of left-behind children in rural areas, and is conducive to promoting these marginalized groups in learning and interpreting those knowledge. Secondly, the negative impact is in the microfocus, dispelling the right image of traditional historical or two-dimensional culture figures, is not conducive to the formation of a clear knowledge construction; in the macro structure, it is not conducive to the formation of a correct view of history or two-dimensional values for left-behind children in rural areas. Because unobservable factors affect the types of children and a series of dependent variables in this article, this paper selects instrumental variables to conduct empirical analysis again. However, there are still limitations in this study. This experiment uses the achievements related to historical knowledge level and two-dimensional culture evaluation of left-behind children as a reference, but the reliability as a single factor needs to be studied. In the future, multiple factors should be used for analysis at the same time. The results of instrumental variable regression of all groups indicate that whether staying or not staying in rural area, children accompanied by parents have no significant difference on their cognitive abilities and emotions. And furthermore, the left-behind children suffered significant impact on their cognitive abilities and emotions, compared with the non-left behind ones. This article adopts the method of horizontal comparison and vertical comparison. Horizontal comparison is to compare the various learning behaviors and academic performance of left-behind children and non-left-behind children in the process of cognitive abilities. Longitudinal comparison is to compare left-behind children within themselves, so as to find out the differences among left-behind children among the problems of left-behind children, so as to adopt different suggestions and policies for different left-behind children when making suggestions and countermeasures.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

This research involving human participants is reviewed and approved by the Professor/Academic Committee of School of Journalism and Communication, Zhengzhou University, PR China. The respondents and their guardians provided their consents with written signatures. All respondents and their guardians are aware of all the personal information related in this study will be kept confidential and used only for the research purpose of this paper. Along with this, they are informed that no personal information or data will be shared with any third party.

AUTHOR CONTRIBUTIONS

ML and JL: conceptualization. MY, MO, and NH: methodology. ML and YX: software. NH, MO, and LA: validation and visualization. JL and YX: formal analysis. MY and MO: data curation. ML: writing—original draft preparation. JL and MO: writing—review and editing. JL and MY: supervision. MY: project administration. JL: funding acquisition. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Zihan Zhang for proofreading the manuscript and also Dr. Hongzhe Kang for the help of collecting data.

REFERENCES

1. Li X, Zhu Q, Li W. An empirical study on the cognitive attitude of rural children using mobile phones (based on the example of children from rural areas in northern Suzhou). *J Russ Media Journal Stud.* (2019) (4):5–23. doi: 10.30547/worldofmedia.4.2019.1
2. Zhu X, Liu J, Liang Z. Are left-behind children more likely to be bullied? Empirical evidence from rural schools in Shaanxi province. *Chin Sociol Rev.* (2020) 52:411–37. doi: 10.1080/21620555.2020.1776603
3. Zhen R, Li L, Liu X, Zhou X. Negative life events, depression, and mobile phone dependency among left-behind adolescents in rural China: an interpersonal perspective. *Child Youth Serv Rev.* (2020) 109:104688. doi: 10.1016/j.childyouth.2019.104688
4. Zhang H. A qualitative study on the rights of marginalized children in Sichuan Province, China. *Child Youth Serv Rev.* (2018) 95:12–8. doi: 10.1016/j.childyouth.2018.10.029
5. Sun T, Tang Q, Liu D, Zhao L, Wang F, Xie H. Mental health literacy about depression among marginalized children in China: a comparative and cross-sectional study. *J Ment Health.* (2021) 30:263–70. doi: 10.1080/09638237.2020.1793117
6. Wang L, Mesman J. Child development in the face of rural-to-urban migration in China: a meta-analytic review. *Perspect Psychol Sci.* (2019) 10:813–31. doi: 10.1177/1745691615600145
7. Jin X, Chen W, Sun IY, Liu L. Physical health, school performance and delinquency: a comparative study of left-behind and non-left-behind children in rural China. *Child Abuse Neglect.* (2020) 109:104707. doi: 10.1016/j.chiabu.2020.104707
8. Minghong W. Research into media exposure behaviors of left-behind children in rural areas from the perspective of media as substitutes. *High Educ Soc Sci.* (2019) 8:24–8. doi: 10.3968/6886
9. Guo J, Chen L, Wang X, Liu Y, Chui CH, He H, et al. The relationship between Internet addiction and depression among migrant children and left-behind children in China. *Cyberpsychol Behav Soc Netw.* (2019) 15:585–90. doi: 10.1089/cyber.2012.0261
10. Tao C, Miller PA, Gao J. Coping with parents' work migration in China: The stress and adjustment of children who got left-behind. In: Szente J, editor. *Assisting Young Children Caught in Disasters.* Berlin, Heidelberg: Springer (2018), p. 57–69. doi: 10.1007/978-3-319-62887-5_6
11. Wang M, Sokol R, Luan H, Perronc BE, Victord BG, Wue S. Mental health service interventions for left-behind children in mainland China: a systematic review of randomized controlled trials. *Child Youth Serv Rev.* (2020) 117:105–14. doi: 10.1016/j.childyouth.2020.105304
12. Fu J, Zhao J. Analysis on the status quo and countermeasures of sports participation of left-behind children in Jilin Province from the perspective of sociology. *Soc Sci Big Data Appl.* (2020) 3:139–43. doi: 10.2991/assehr.k.201030.029
13. Li J, Chen Y, Zhang J, Lv MM, Välimäki M, Li YF, et al. The mediating role of resilience and self-esteem between life events and coping styles among marginalized adolescents in China: a cross-sectional study. *Front Psychiatry.* (2020) 11:1190. doi: 10.3389/fpsyt.2020.560556
14. Niu G, Chai H, Li Z, Wu L, Sun X, Zhou Z. Online parent-child communication and left-behind Children's subjective well-being: the effects of parent-child relationship and gratitude. *Child Indic Res.* (2020) 13:967–80. doi: 10.1007/s12187-019-09657-z
15. Li X, Zhu Z S, Tao Z Y, et al. The effect of return migration on rural children's cognitive ability in poverty-stricken areas: an empirical study based on 137 rural boarding primary schools. *China Rural Econ.* (2019) 9:11–3.
16. Liu Z, Zhou B. The lasting impact of parental migration on children's behavioral outcomes: evidence from China. *Chin Sociol Rev.* (2020) 52:438–61. doi: 10.1080/21620555.2020.1776604
17. Yue Z, Liang Z, Wang Q, Chen X. The impact of parental migration on depression of children: New evidence from rural China. *Chin Sociol Rev.* (2020) 52:364–88. doi: 10.1080/21620555.2020.1776601
18. Coleman R, Chen T. Forgotten frames: proposing the concept of "digressive framing" using left-out frames in Chinese media coverage of left-behind children. *Int J Commun.* (2020) 7:14–21.
19. Chen Y, Zhu L, Chen Z. Family income affects children's altruistic behavior in the dictator game. *PLoS ONE.* (2019) 8:19–22. doi: 10.1371/journal.pone.0080419
20. Lam T, Yeoh BSA, Hoang LA. *Transnational migration and changing care arrangements for left-behind children in Southeast Asia: A selective literature review in relation to the CHAMPSEA study. Asia Research Institute Working Paper, No. 207.* (2019). Available online at: https://ari.nus.edu.sg/wp-content/uploads/2018/10/wps13_207.pdf
21. Chen X, Jiang X. Are grandparents better caretakers? Parental migration, caretaking arrangements, children's self-control, and delinquency in rural China. *Crime Delinq.* (2019) 65:1123–48. doi: 10.1177/0011128718788051
22. Fan C, Yan D, Xio F, Li A, An J, Kang X. Advanced data analytics for enhancing building performances: from data-driven to big data-driven approaches. *Build Simul.* (2021) 14:3–24. doi: 10.1007/s12273-020-0723-1
23. Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: a review. *J Manuf Syst.* (2021) 62:738–52. doi: 10.1016/j.jmsy.2021.03.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Li, Yasin, Osman, Hashim, Ang and Xue. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fpubh.2022.896967

Digital Analysis of Smart Registration Methods for Magnetic Resonance Images in Public Healthcare

Tao Chen[1], *Mengxue Yuan*[2], *Jiajie Tang*[2*] *and Long Lu*[2,3*]

[1] School of Information Technology, Shangqiu Normal University, Shangqiu, China, [2] School of Information Management, Wuhan University, Wuhan, China, [3] Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

Brain development and atrophy accompany people's life. Brain development diseases, such as autism and Alzheimer's disease, affect a large part of the population. Analyzing brain development is very important in public healthcare, and image registration is essential in medical brain image analysis. Many previous studies investigate registration accuracy by the "ground truth" dataset, marker-based similarity calculation, and expert check to find the best registration algorithms. But the evaluation of image registration technology only at the accuracy level is not comprehensive. Here, we compare the performance of three publicly available registration techniques in brain magnetic resonance imaging (MRI) analysis based on some key features widely used in previous MRI studies for classification and detection tasks. According to the analysis results, SPM12 has a stable speed and success rate, and it always works as a guiding tool for newcomers to medical image analysis. It can preserve maximum contrast information, which will facilitate studies such as tumor diagnosis. FSL is a mature and widely applicable toolkit for users, with a relatively stable success rate and good performance. It has complete functions and its function-based integrated toolbox can meet the requirements of different researchers. AFNI is a flexible and complex tool that is more suitable for professional researchers. It retains most details in medical image analysis, which makes it useful in fine-grained analysis such as volume estimation. Our study provides a new idea for comparing registration tools, where tool selection strategy mainly depends on the research task in which the selected tool can leverage its unique advantages.

Keywords: MRI, registration technology, feature comparison, SPM, FSL, AFNI

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Irfan Uddin,
Kohat University of Science and Technology, Pakistan
Achyut Shankar,
Amity University, India
Xin Qi,
Waseda University, Japan

****Correspondence:***
Jiajie Tang
jiajietang@whu.edu.cn
Long Lu
bioinfo@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *15 March 2022*
Accepted: *15 April 2022*
Published: *06 June 2022*

Citation:
Chen T, Yuan M, Tang J and Lu L (2022) Digital Analysis of Smart Registration Methods for Magnetic Resonance Images in Public Healthcare. Front. Public Health 10:896967. doi: 10.3389/fpubh.2022.896967

INTRODUCTION

Image registration plays an essential role in image fusion, pattern recognition, and voxel-wise group analysis. Establishing correspondences across brains for comparison and group analysis is almost universally done by registering images to one another directly or *via* a template. Image registration is generally categorized into two groups, rigid and non-rigid registration. Rigid registration only consists of rotation and translation (1). Non-rigid registration is more complex and would take deformation methods, such as affine transformation and spline transform. There are a large number of optimized non-rigid registration algorithms and available registration tools, including

advanced normalization tools (ANTS), analysis of functional neuroimages (AFNI), automated image registration (AIR), Drop, FMRIB's software library (FSL), and statistical parametric mapping (SPM). Thus, a comprehensive evaluation of different registration methods has become a research topic of interest. It is the basis for users to choose the most suitable methods for the specific research problem and for algorithm developers to be better informed theoretically (2). However, it is hard to assess image registration algorithms under different research contexts accurately. According to literature and survey, the main obstacles to building a unified evaluation standard of registration algorithm are as follows:

(1) No ground truth results are supplied in practice. In reality, the registration results are mainly assessed by an experienced expert. The lack of golden truth makes the evaluation research less convincing in terms of data. Also, it is hard to be flexible and straightforward to evaluate the performance of registration methods in a complex reality.
(2) Various parameter configurations for registration algorithms may lead to registered images of different quality. Experienced investigators can set appropriate parameters based on images' information. For the same input image, the difference in parameter setting would cause different registered outcomes (3).
(3) The same image data may be processed with different registration techniques to complete various tasks. Therefore, the data sets should be classified first and then preprocessed by other data registration methods according to different data characteristics.
(4) Data quality and raw data quality substantially impact registration results, which means that the data received by the registration program are quite different. The same algorithm may make a profound difference in the registration outcome when dealing with data with a significant quality difference.

Some researchers have done relevant research about the evaluation and comparison of registration algorithms to investigate this topic. Hellier et al. compared five different fully automated non-linear brain image registration software programs using the quantitative measures (4–6). Klein et al. evaluated 14 non-linear deformation algorithms applied to human brain MRI registration on four datasets and ranked these algorithms (7). Rajagopalan and Pioro investigated disparate voxel-based morphometry (VBM) results between SPM and FSL software in patients with ALS to determine which tool has the best performance of VBM in the ALS disease setting (8). Most recently, Dadar et al. compared publicly available linear MRI stereotaxic registration techniques by viewing the registered images by an expert rater to assess the quality of the registrations (9). At the same time, Viergever et al. made a retrospective view on the past two decades of medical image registration. They mentioned that "validation of registration methods and translation of image registration research results to the clinical practice still is the highlighted research and be more urgent than two decades ago (10)." Meanwhile, Rohlfing provides experimental evidence that registration accuracy measures such as tissue label overlap scores, image similarity, image difference, or transformation inverse consistency error, even when used in combination, cannot distinguish accurate from inaccurate registrations (11). Some previous accuracy standards are no more extended evaluation standards. Simulated data and a database with expert landmark annotations have been employed for comparison in the last decade to measure accuracy.

The registration of brain MRI with common templates is a long-standing problem. Different individuals' brain shapes and cortical topology are very different, especially in diseases affecting brain morphology and structure, such as brain atrophy. Previous researches mainly focus on registration accuracies, such as overlap, volume similarity, and registration error. However, the choice of registration technology has a more significant impact on image registration results. Few studies have used different registration tools to process image features. This means that the existing studies on the selection of image registration technology are insufficient. The image registration methods used in many studies are not standardized, and the accuracy level after registration is unknown and difficult to be quantified. Such problems will significantly affect the results of follow-up research. Therefore, the evaluation of image registration technology only at the accuracy level is not comprehensive. When facing different image processing projects, researchers need a flexible selection scheme to help them select the appropriate registration technology according to data sets' characteristics and personalized needs. Here, we investigate the performance of three publicly registration tools in two typical different disease cases, aiming to find a suitable registration tool for different disease types and research focus. Our research uses ABIDE and TCGA date sets to conduct experiments and deeply compares the registration performance of FSL, SPM12, and AFNI. In addition, we compare the three mainstream image registration technologies in terms of user experience and recommend them to different groups.

Our research investigates the registered results in two disease types, autism spectrum disorder (ASD), and glioma. Glioma may cause deformation in the morphology of the brain, and physicians can use high-resolution structural MRI images to detect the focus of infection and glioma grades. By contrast, ASD is connected with brain function and emotion and is often a pervasive developmental disorder. The pathogenesis of ASD is related to heredity, so the patient would have ASD at a very young age. The structural MRI and functional MRI images can be used to diagnose ASD in the clinic. We collect the MRI image data from public databases to construct our disease dataset. We do not intend to evaluate accuracy and error by manual labeling and check which are not available in reality. We plan to compare the features and characteristics of different tools' results because registration techniques are designed for follow-up analysis and application. We focus on the difference in registered results under various analyses, such as texture and edge-gradient features, voxel-based morphometry, and the impacts of this inter-method discrepancy on prediction and detection. From the application's perspective, we investigate the difference between three mature and powerful software to provide instructions for

TABLE 1 | Dataset information.

Dataset	ABIDE II-ETH	TCGA-LGG
Slice thickness (mm)	0.9	2.5
No. of slices	180	69–80
Scan matrix	256 ×256	256 × 256
Repetition time (ms)	8.4	10,002
Echo time(ms)	22	125
No. of scans	31	27

different researchers to select a suitable tool and more quickly get better achievements.

The structure of our study is as follows: in Materials and Methods, we describe the data acquisition and registration tools, then the strategy for data processing, and the ABIDE II and TCGA-LGG data set used to assess different registration tools. We focus on the difference in registered results under various analyses, such as texture and edge-gradient features, voxel-based morphometry, and the impacts of these inter-method discrepancies on prediction and detection. As a result, we instruct different researchers to select a suitable tool and more quickly get better achievements by investigating the difference between three mature and powerful software. Finally, in the discussion, we summarize the current work, emphasize the importance of selecting appropriate registration methods, and discuss the limitations and future improvement ideas.

MATERIALS AND METHODS

Data Acquisition and Quality Control

This section introduces the basic information of ABIDE II-ETH and TCGA-LGG. **Table 1** presents the basic information of the database.

(1) ABIDE II-ETH. Autism Brain Imaging Data Exchange II was established to further promote discovery science on the brain connectome in ASD (12). It is a multi-center and multi-scanner study involving 19 sites, ten charter institutions and seven new members, overall donating 1,114 datasets from 521 individuals with ASD and 593 controls (age range: 5–64 years). We select the ETH dataset to finish our research. There are 37 samples in the ETH set (age range 14–31 years), 13 for ASD and 24 for typical control (TC). Then, we manually check the ETH dataset to discard the bad quality image with serious artifacts.

(2) TCGA-LGG. The Cancer Genome Atlas Low-Grade Glioma (TCGA-LGG) data collection is part of a more significant effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA) (13). It contains CT and MRI data collected from 199 patients, and the number of studies in TCGA-LGG is 224. To gain the low-level glioma data with the same scan sequence and quality similar to the autism data set, we manually select a small part of the subject containing T2-w and T2-Flair sequence MRI data to finish our research.

Registration Tools

SPM12

The SPM has several versions, and the latest is the SPM12 (14, 15). SPM12 is designed to work with MATLAB to run on Windows systems, while other similar tools are always supported by Linux or Mac and are easy to install. If someone is new to imaging, SPM12 is a recommended choice because plenty of tutorials give practical instructions on how to implement the various methodologies.

The spatial normalization method in SPM12 is a unified model based on a probabilistic approach. It combines the functions of image registration, tissue classification, and bias correction in the same generative model. The model is based on a Gaussian mixture and is extended to incorporate a smooth intensity variation and non-linear registration with tissue probability maps (15).

FSL

FSL is a software package developed by the Oxford Center for Functional MRI of the Brain (Oxford University), composed of image analysis and statistical tools for neuroimage data study. Flirt or Fnirt performed image registration in FSL. Flirt uses a multi-start, multiresolution global optimization method (16, 17), tailored explicitly for volumetric registration of brain images to give a reliable estimate of the worldwide minimum given some time restriction. The optimization algorithm uses four resolution levels: $n = 8, 4, 2, 1$ mm. Initially, the procedure starts with a large-scale search at 8 mm resolution (e.g., applying a set of initial rotations). Following this, various local optimizations are performed with multiple starting points in the local neighborhood of the best issues identified in the search. Then a series of multi-start local optimizations at 4 mm resolution is completed.

AFNI

AFNI is a leading software suite of C, Python, R programs, and shell scripts primarily developed to analyze and display anatomical and functional MRI (fMRI) data (18). The software is made to run on virtually a Unix system with X11 and Motif displays and is a widely used tool for image preprocessing. There are plenty of functions to perform registration, e.g., @auto_tlrc, 3dWarp, 3dWarpDrive, 3dAllineate, @toMNI_Awarp, @toMNI_Qwarp. The common function for registration is @auto_tlrc. This function is run by default parameters. Here, we use @auto_tlrc to execute registration (19).

Data Processing

We download data from the website and also check the quality and image label. After eliminating unqualified data, there are 33 for autism data sets (ASD + TC) and 45 for glioma data sets (glioma + TC) as a study data source. It is learned from **Table 2** that the available data formats for each tool are not entirely the same. Firstly, we convert images of DICOM format

TABLE 2 | Format requirements.

Public tools data format	DICOM	ANALYZE	NIFTI_PAIR	NIfTI-1	NIFTI_GZ	HEAD/BRIK	MINC	CTF	ASCII
SPM	√	√	√	√					
FSL		√	√	√	√				
AFNI	√	√	√	√	√	√	√	√	√

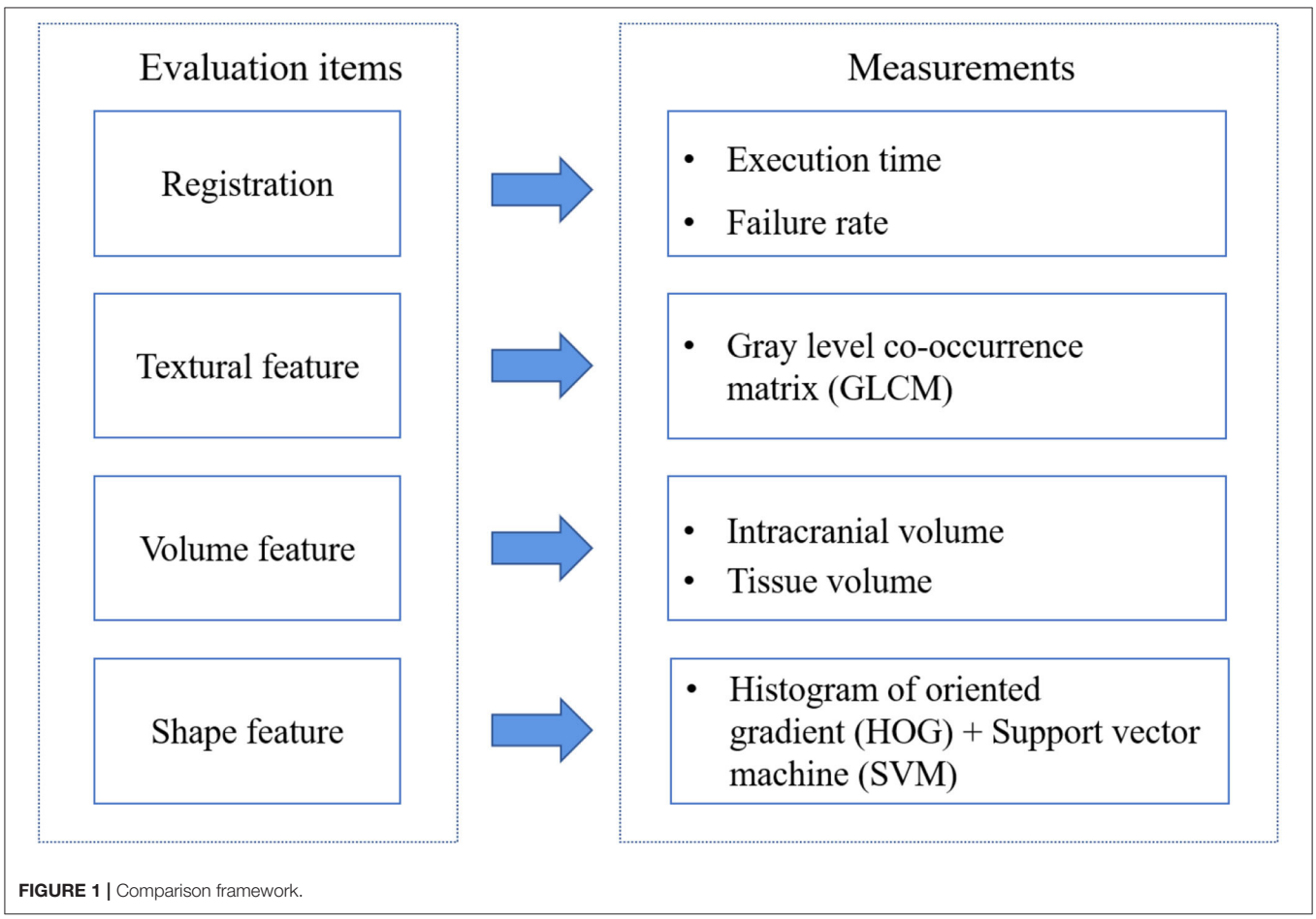


FIGURE 1 | Comparison framework.

into NIFTI-1 format. NIFTI-1 format is also the most commonly used international MRI format.

To explore the features difference among the three tools registered results, we construct the comparison framework shown in **Figure 1**. After organizing the dataset, we execute the registration step by AFNI, SPM, and FSL. T1-w and T2-FLAIR images were all registered to the standard space. SPM12 operations are finished on MATLAB R2012a. FSL and AFNI are installed and run in Ubuntu14 on a virtual machine.

There are three primary features for images: color, texture, and shape. Here, MRI images are gray images without color features. We consider volume as one of the evaluation items. Then we will introduce each item and its measurements.

Textural Feature

A textural feature is a diagnostic tool for physicians when reading MRI images. Learning from this, Zacharaki et al. extract MRI texture and shape features, constructing the feature vector to train a support vector machine (SVM) classifier to classify brain tumor type and grade (20). A textural feature is usually employed for tumor grading in tumor studies because it provides fine-grained, repeatable information. Besides, for autism, commonly diagnosed by behavioral, chemical, clinical, structural, and functional changes in the brain, MRI texture represents a new image feature, which can supplement the traditional image features (such as volume measurement). Vidhusha and Anandhan focused on extracting texture features for autistic and control subjects, and validated them using neural classifiers to achieve automatic detection of autism (21). Textural features are extracted from original MRI ROI images rather than registered results, but here, we explore the textural difference between results suffering different processing pipelines, and we will do a single statistical analysis to discuss the significant difference.

The textural feature has several extraction methods, including statistical, model, signal processing, and geometric forms. Here, we choose the gray level co-occurrence matrix (GLCM) proposed by Haralick et al. regarded as one of the most promising texture analysis methods (22, 23). It estimates the image properties related to second-order statistics. GLCM is a joint distribution describing the gray level of two pixels with a particular spatial positional relationship. Five derived variables were calculated from the GLCM matrix, contrast, correlation, energy, homogeneity, and IDM to investigate the difference between textural features. Then these computations will be described in detail next.

Contrast. The deeper the texture groove, the greater the contrast and the clearer the visual effect; conversely, the contrast is small, the groove is shallow, and the effect is blurred.

$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i,j) \tag{1}$$

Correlation. It represents the consistency of the image texture. If there is a horizontal direction texture in the image, the COR of the horizontal direction matrix is larger than the COR value of the remaining matrix.

$$correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij - \mu_x \mu_y}{\sigma_x(i)\sigma_y(j)} \tag{2}$$

Energy. It is the sum of the squares of the gray level co-occurrence matrix values, so it is also called angular second moment, which reflects the uniformity of image gray distribution and texture thickness. A large energy value indicates a textured pattern that is more uniform and regular.

$$energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i,j))^2 \tag{3}$$

Homogeneity. It reflects the density of elements in GLCM relative to the diagonal distribution of GLCM.

$$homogeneity\,1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1+|i-j|} \tag{4}$$

Inverse Difference Moment. IDM measures the local homogeneity of an image and reports the inverse difference moment of an image. IDM weights are the inverse of contrast weights. It has the value that determines whether the print is textured or non-textured.

$$IDM = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+k^2} \tag{5}$$

Volume Feature

Volume could provide significant information on pathological study and can capture volume change in the brain. Quantitatively detecting the density and volume of brain tissue at the voxel level can reflect the differences in brain tissue components and characteristics in different brain regions. The latest study explores that the inter-method discrepancies in brain volume estimation may drive inconsistent findings in autism (3). In this part, we employ SPM12 as the brain volume estimation tool. The images are segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) volume in the native space to estimate the volume by using the New Segment tool of SPM12 (6). Spm_get_volumes script is used to calculate the tissue volumes using c1, c2, and c3 images corresponding to native space tissue maps of GM, WM, and CSF, respectively. Native space volumes are selected to minimize volume changes due to spatial transformations. Total intracranial volume (TIV) is calculated as the sum of the GM, WM, and CSF volumes in the native space of the structural MRI images.

Shape Feature

We calculate textural matrix to understand the textural feature and compare the brain tissue volume estimation, but it is hard to compare shape differences for shape features. Thus, we employ machine learning algorithms using shape features as the input and record the prediction accuracy to evaluate the shape feature performance. Based on the shape features extracted from MRI, machine learning algorithms could find the standard classifier function to distinguish the healthy and disease cases. To capture shape features in MRI images, we extracted the histogram of oriented gradient [HOG; (24)], a prominent local image feature widely used in computer vision applications, from MRI images to establish the feature sets. The HOG technique counts occurrences of gradient orientation in localized portions of an image. When studying the brain tumor, the HOG could describe local object appearance and shape within an image by distributing intensity gradients or edge directions. Thus, this descriptor can detect the abnormal edge of the tumor issues and help physicians find the lesions. About ASD, the HOG descriptor may also find some markers to distinguish the healthy and disease cases.

Then, the classification method is the support-vector machine (SVM), the supervised learning model. The combination of HOG and SVM is the classical strategy in object detection, especially in human detection (24). In the model, the cross-validation (CV) method is used to evaluate the performance of classifiers trained by different MRI data sets registered by three tools. The widely used methods in brain image analysis include leave-one-out cross-validation (25, 26), leave-two-out cross-validation (27–29), k-fold cross-validation (1), and stratified k-fold cross-validation (30, 31). The cross-validation results are conducted from the strategy that divides the raw data into K groups (K-fold), and each subset of data is used as a verification set. The remaining K-1 subset data is used as a training set. Then the K models, respectively, evaluate the results in the verification set, which effectively utilizes limited data, and the results are as close as possible to the model's performance on the test set. According

to review studies and previous research, we use stratified ten-fold cross-validation to measure classification performance. All operations are performed in python3.7 in Ubuntu14 on a virtual machine.

RESULTS

Time and Failure Rate

Execution time is a crucial factor that needs to consider, which can reflect the effectiveness of tools. During the registration step, we calculate the time in each registration pipeline. **Table 3** and **Figure 2** show that AFNI gets the best speed performance and then SPM12 while FSL is the slowest.

Registration is also regarded as an optimization process, which means there are failures in this step. We also collect the failure rate in the registration step. From **Table 3**, the failure rate in AFNI is higher than in FSL, and in FSL it is higher than in SPM. Auto-registration in SPM12 is based on prior knowledge using the tissue probability map to find the optimum transformation to achieve high and accurate registration. What is more, in SPM12, there is no feedback mechanism, and the final transformed results will be returned, whether good or bad. By contrast, FSL and AFNI both have the mechanism that if the cost function cannot be optimized, they will return the failure message and some additional suggestions to help the user optimize the registration command and get better results. So, AFNI and FSL have improvable abilities while SPM12 is poor.

The purpose of our study is not to attempt to find the most accurate one that has been investigated in many previous studies (7, 9, 32–34). We compare the registered results in the image to explore the degree of change in the features and structures made by the three registration tools and find the difference between these changes to help the user select the registration tool based on their registration results difference and research purpose.

TABLE 3 | Time acquisition and failure rate.

	ABIDE II-ETH		TCGA-LGG	
	Time (Average)	Failure rate (%)	Time (Average)	Failure rate (%)
SPM12	94 s	0	100 s	0
FSL	110 s	0	4 min	16
AFNI	78 s	0	86 s	20

Texture Feature

We extract valuable features from the registered images by different tools to find the difference. Brain regions were extracted using the brain extraction tool [BET; (35)]. We extracted textural features from disease data sets to study the difference between registration tools and disease types (36, 37). Input images are brain-extracted MRI images to gain whole-brain texture GLCM matrix and calculate the five features, i.e., contrast, correlation, energy, homogeneity, and inverse difference moment (IDM).

Figure 3 shows the results of the comparison of texture features. The value distribution is similar among SPM, FSL, and AFNI and shows no significant difference. But FSL gets a higher value, implying that FSL is more inclined to retain more uniform and regular texture patterns. AFNI has a higher correlation value corresponding to its biggest size of MRI holding more pixels and details that could be measured more similar pattern.

In the statistical analysis, SPM12 values shows significant difference with AFNI ($p = 0.017314065 < 0.05$, 95%) and FSL ($p = 0.046368617 < 0.05$, 95%), indicating that SPM12 registered images have more sharp textural changes. A larger contrast value correlates with a greater disparity in intensity values among neighboring voxels, suggesting that SPM12 processed results have more sharp intensity changes and deeper texture grooves. Next, the inverse difference moment, also known as

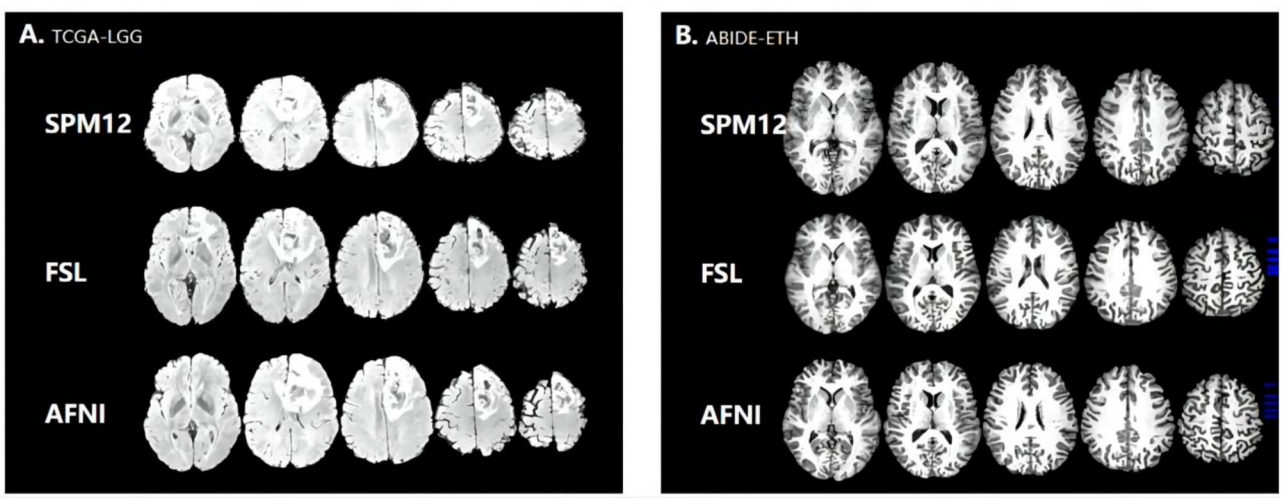


FIGURE 2 | Registered results by three tools. **(A)** TCGA-LGG. **(B)** ABIDE-ETH.

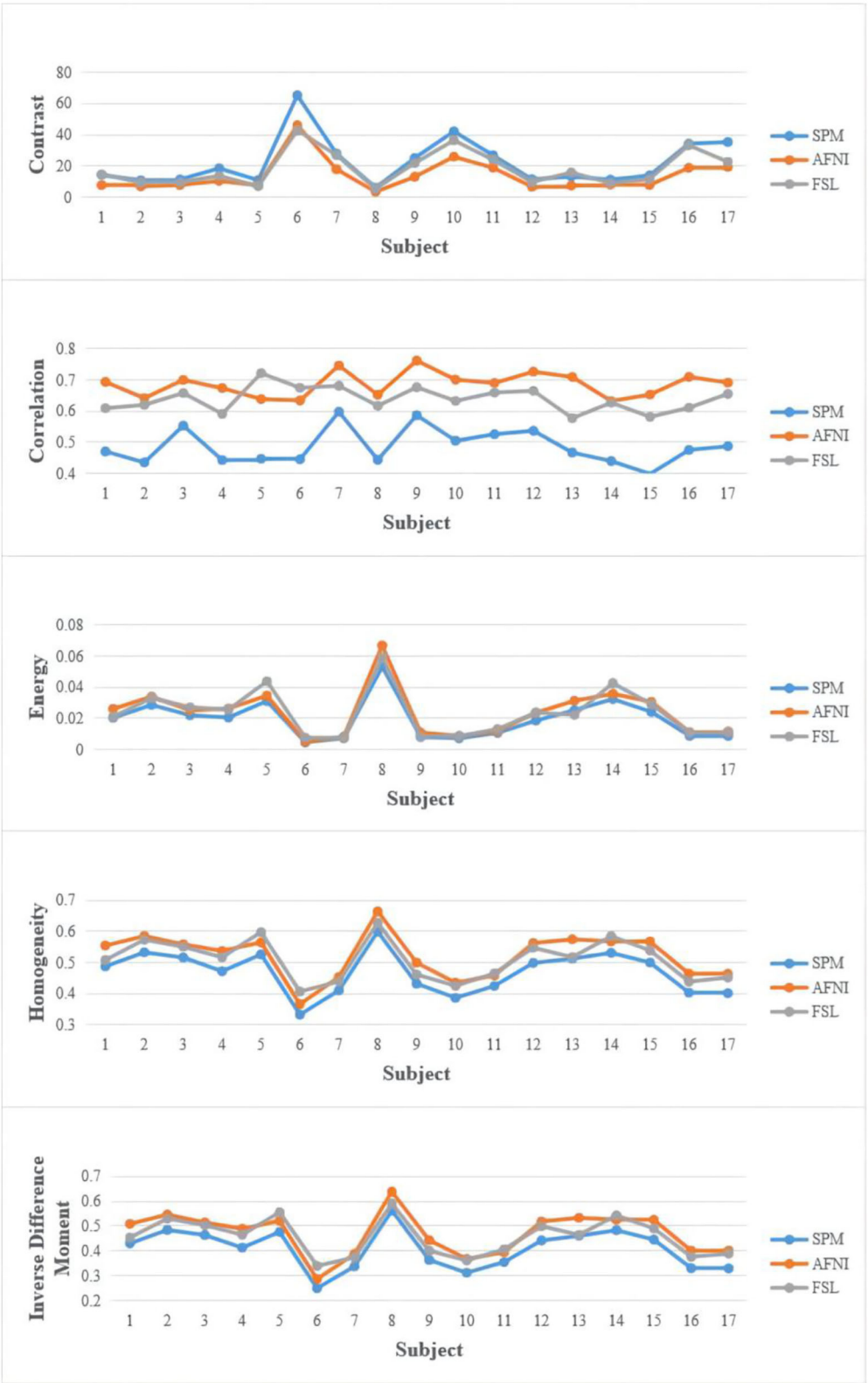


FIGURE 3 | Texture feature extracted from the dataset.

TABLE 4 | Brain volume estimation results.

	TIV (mean)	GM% (mean)	WM% (mean)	CSF% (mean)
TCGA-LGG	1.5364	39.89	45.17	14.94
SPM (TCGA)	1.9012	35.41	41.61	22.97
AFNI (TCGA)	1.8936	36.08	45.13	18.79
FSL (TCGA)	1.5123	33.62	43.89	22.49
ETH	1.2960	58.79	32	9.21
SPM (ETH)	1.8776	54.21	30.17	15.02
AFNI (ETH)	1.8977	55.63	29.63	14.74
FSL (ETH)	1.5641	55.81	28.01	16.19

ETH = ABIDE II-ETH.

the inverse variance, reflects the clarity and regularity of texture. The images with clear, regular, easy to describe texture would have a larger value demonstrating that these images are more in line with human feelings. The AFNI is clearer than the other two data sets in **Figure 2**, which is consistent with our calculated values. SPM12 also has significant difference with AFNI ($p = 0.018590934 < 0.05$, 95%) and FSL ($p = 0.044412676 < 0.05$, 95%) as well. We guess the reason is that AFNI retains more detailed information, including uniform and sharp textural change, while SPM12 abandons these details and hold sharp textural feature.

Brain Volume Estimation

Some brain diseases may give rise to the volume change in the brain, such as edema, which is obvious in the image. We calculate the TIV and the percentage of gray matter, white matter, and CSF in the original and registered structural MRI images. We set the volume estimation from the MRI in the original space as the standard result. From **Table 4** and **Figure 4**, AFNI shows the closest results with similar GM, WM, and CSF percentages, and sometimes overestimates the total volume. The whole-brain volume estimated by SPM was always higher than the standard TIV. FSL estimates of CSF are the highest.

Shape Feature-Based Classifier

Finally, we compare the evaluated accuracy of the cross-validation. For the ASD dataset, the number of healthy and disease cases is 22 and 11, respectively. For the glioma dataset, the number of healthy and disease cases is 22 and 17, respectively, and the k that we use in k-fold is 10.

From **Table 5**, we can observe the difference between the registration tools. The performance of SPM shows the highest accuracy, while AFNI is better than FSL. Since the HOG feature represents the shape information by edge directions, which collects contrast information, this result is consistent to compare results found in the texture feature part. MRI images registered by SPM retain richer contrast sources, which may mean that SPM is a suitable preprocessing tool when dealing with structural problems. The SPM performs best in our structural HOG feature model.

DISCUSSION

To estimate the performance of registration algorithms of SPM, FSL, and AFNI, we calculate the registration time, compare the texture feature extracted from skull-striped structural MRI images, and investigate the inter-methods discrepancies in brain volume estimate. Finally, we compare these registered images' applied performance in the classification model and we get the conclusion that SPM12 is usable for newcomers or employed when dealing with basic and normal processing workflow because it is like a black box that users cannot know the registration process details. FSL is a powerful toolkit for research. Its function-based integrated toolbox facilitates researchers for the different hierarchy of requirements. However, FSL's semi-integrated design cannot meet scholars' requirements for function module combinations and workflow model optimization. In contrast, AFNI is cumbersome and complex to use. However, it can provide some personalized customization and allow some threshold adjustments, which is very useful for many professional researchers, but it may be difficult for novices to use.

According to our analysis conclusion, tool selection mainly depends on research tasks. If the focus of studies is only MRI images' single feature, it is wise to choose the feature-targeted tools, for example, the contrast feature of texture feature for SPM, and if studies require multiple features or detailed features, AFNI will be a good choice. When involving multi-modal MRI sequences, like DWT, PET, and DTI, FSL would become the first choice. For glioma and autism spectrum disorder, the glioma is more sensitive to contrast features, so SPM12 may perform better registration results when studying the glioma dataset. When using the shape features like HOG, SIFT, or LBA, SPM12 may have the best performance on comparing AFNI and FSL. Tumor detection, prediction, and classification are likely to employ these structural and edge-related features, and so SPM12 will be an advisable process tool with stable performance and feature-sensitive advantages, facilitating tumor study.

Image registration research is still progressing steadily, and new research results are emerging. At present, there is still no unified evaluation criterion. Experience and habits are still the main factors affecting people's choice of methods and tools. Our study provides a new idea for the comparison of registration tools. Using data sets from different diseases, we study the differences in texture features, structure, and classification performance of the three registration tools, discuss their characteristics, and analyze the causal correlation between features and classification models. Our study is not without limitations. The focus of this study is only on structural images, while the registration between functional images and other modal data has not been investigated in detail. Secondly, only three kinds of public registration tools are selected, and some tools in current research are omitted. In addition, the default parameter settings are adopted in the study. Still, in the specific analysis, different parameter settings will significantly impact the results, and the appropriate parameter configuration can optimize the processing results. There is some research on MRI registration

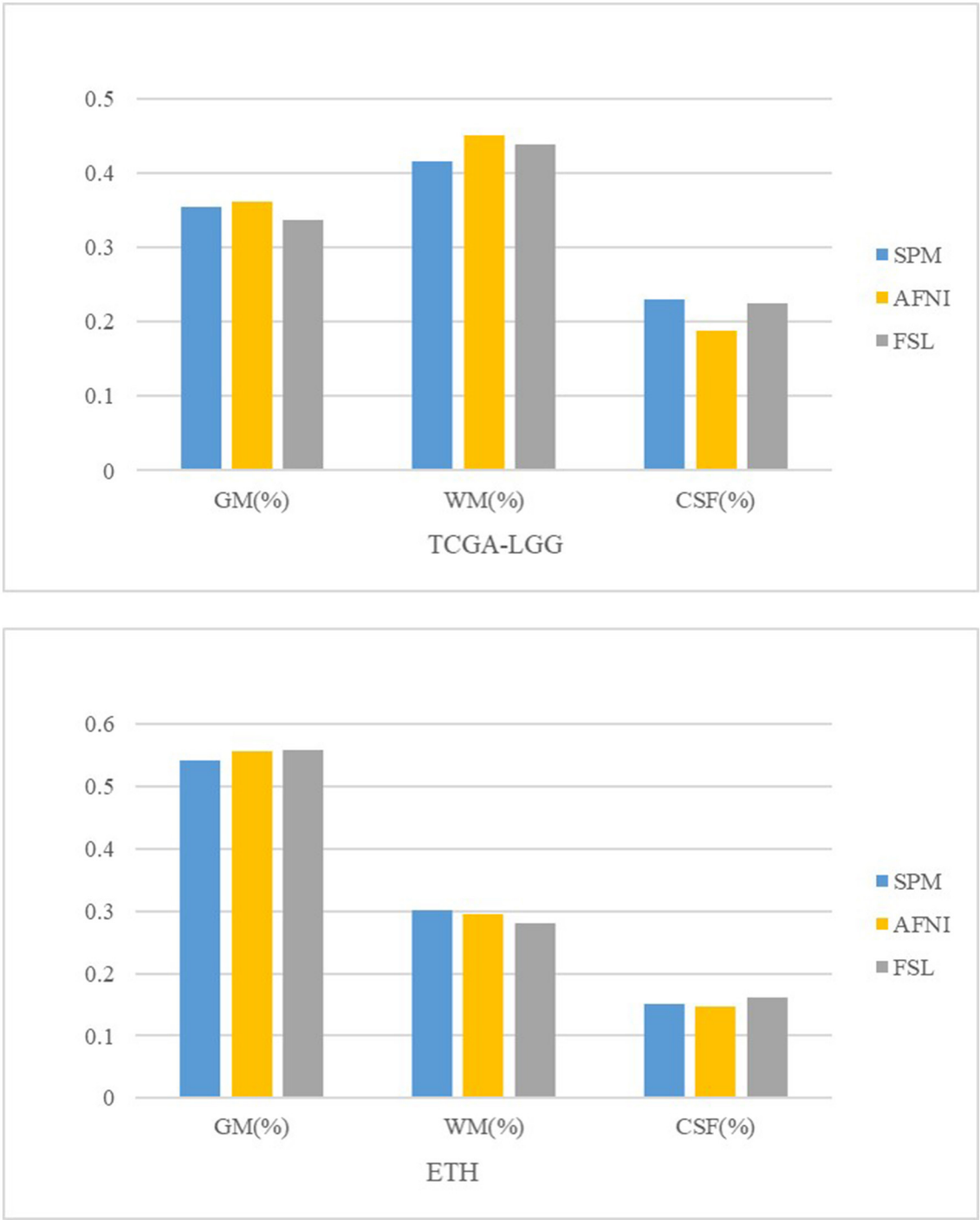


FIGURE 4 | Brain tissue volume comparison.

using a deep learning algorithm, which may achieve better performance than traditional MRI registration methods. In the next step, we need to compare and analyze the registration results of these methods (38, 39).

CONCLUSIONS

Although SPM has the most stable performance, it also produces poor registered results when dealing with low-quality data

TABLE 5 | Prediction results on two datasets.

ABIDE II-ETH dataset	SPM %	AFNI %	FSL %
Accuracy (high)	78.79	72.73	63.64
Accuracy (low)	69.70	63.64	54.55
TCGA-LGG dataset	**SPM %**	**AFNI %**	**FSL %**
Accuracy (high)	92.31	88.83	81.17
Accuracy (low)	86.24	80.52	72.08

sets. According to its advantages and drawbacks, SPM12 has a stable speed and success rate but does not guarantee the processed results, especially when handling low-quality images. FSL is a mature and widely applicable toolkit for employers, with relatively stable performance, a success rate, and complete functions. Its graphical user interface provides basic processing flow operation, while multi-parameter settings and failure feedback mechanisms can realize complex user processing and meet the optimization requirements. It is an excellent toolbox for beginners and proficient hands. Last, AFNI, with quite a fast speed and complex modules, can meet various requirements. But it is not friendly to newcomers, even though it offers an integrated processing function-similar function and modules, unique data format, and command-line operation, all confusing to beginners.

The texture feature is significant in image registration. SPM registration enhances the contrast feature of the texture feature, giving up other detailed information. FSL registration contains enough contrast information and most of the details. AFNI registration is apparent to view details in the brain structure. Due to plentiful details, too much information may make finding key features to differentiate brain structure by single textual features. Therefore, SPM registered image is easy to distinguish the different and distinct tissue types in the images. MRI images processed by AFNI and FSL are more suitable for complex and fine-grained investigations.

In conclusion, AFNI will be a good product application for its immediate and quick response with processed results but requires extra steps to ensure quality. It also plays an essential role in academics and research with its excellent flexibility, which could quickly test various combined processing steps in research. It is also the only MRI registration tool that supports all formats. SPM12 always works as a guide tool for newcomers to medical imaging. A rich online teaching resource and simple and easy-to-learn operations make everyone accomplish the basic processing flow and get stable results. Its support for the Windows environment is the main reason for its popularity. FSL combined advantages of both, providing GUI for basic and simple processing flow and customized parameters setting for flexible and optimized procedure. In addition, FSL is a suitable tool for other MRI sequences like DWT and has been widely used in numerous studies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://www.nitrc.org/frs/downloadlink.php/9090 ABIDE II-ETH (39); https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG TCGA-LGG (224).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This research was funded by National Natural Science Foundation of China (61772375, 61936013, and 71921002), the National Social Science Fund of China (18ZDA325), National Key R&D Program of China (2019YFC0120003), Natural Science Foundation of Hubei Province of China (2019CFA025), Independent Research Project of School of Information Management of Wuhan University (413100032), the Key R&D and Promotion Projects of Henan Province of China (212102210521), and the Key Scientific Research Project in Colleges and Universities of Henan Province of China (22A520038).

REFERENCES

1. Zhe L, Deng D, Guang-Zhi W. Accuracy validation for medical image registration algorithms: a review. *Chin Med Sci J.* (2012) 27:176–81. doi: 10.1016/S1001-9294(14)60052-3
2. Ou Y, Akbari H, Bilello M, Da X, Davatzikos C. Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights. *IEEE Trans Med Imaging.* (2014) 33:2039–65. doi: 10.1109/TMI.2014.2330355
3. Katuwal GJ, Baum SA, Cahill ND, Dougherty CC, Evans E, Evans DW, et al. Inter-Method discrepancies in brain volume estimation may drive inconsistent findings in autism. *Front Neurosci.* (2016) 10:439. doi: 10.3389/fnins.2016.00439
4. Hellier P, Barillot C, Corouge I, Gibaud B, Goualher GL, Collins DL, et al. Retrospective evaluation of inter-subject brain registration. In: *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention.* Utrecht: Springer-Verlag (2001). p. 258–65. doi: 10.1007/3-540-45468-3_31
5. Hellier P, Barillot C, Corouge I, Gibaud B, Le Goualher G, Collins DL, et al. Retrospective evaluation of intersubject brain registration. *IEEE Trans Med Imaging.* (2003) 22:1120–30. doi: 10.1109/TMI.2003.816961
6. Hellier P, Ashburner J, Corouge I, Barillot C, Friston KJ. Inter subject registration of functional and anatomical data using Spm. In: *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II.* Springer-Verlag (2002). p. 590–7. doi: 10.1007/3-540-45787-9_74
7. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain Mri registration. *NeuroImage.* (2009) 46:786–802. doi: 10.1016/j.neuroimage.2008.12.037
8. Rajagopalan V, Pioro EP. Disparate voxel based morphometry (Vbm) results between Spm and Fsl Softwares in Als patients with frontotemporal dementia: which Vbm results to consider? *BMC Neurol.* (2015) 15:32. doi: 10.1186/s12883-015-0274-8
9. Dadar M, Fonov VS, Collins DL. A comparison of publicly available linear Mri stereotaxic registration techniques. *NeuroImage.* (2018) 174:191–200. doi: 10.1016/j.neuroimage.2018.03.025

10. Viergever MA, Maintz JBA, Klein S, Murphy K, Staring M, Pluim JPW. A survey of medical image registration - under review. *Med Image Anal.* (2016) 33:140–4. doi: 10.1016/j.media.2016.06.030
11. Rohlfing T. Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE Trans Med Imaging.* Pune (2012) 31:153–63. doi: 10.1109/TMI.2011.2163944
12. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry.* (2014) 19:659–67. doi: 10.1038/mp.2013.78
13. Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (Tcga): an immeasurable source of knowledge. *Contemp Oncol.* (2015) 19:A68–77. doi: 10.5114/wo.2014.47136
14. Ashburner J, Neelin P, Collins DL, Evans A, Friston K. Incorporating prior knowledge into image registration. *NeuroImage.* (1997) 6:344–52. doi: 10.1006/nimg.1997.0299
15. Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Map.* (1999) 7:254–66. doi: 10.1002/(SICI)1097-0193(1999)7:4<254::AID-HBM4>3.0.CO;2-G
16. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage.* (2002) 17:825–41. doi: 10.1006/nimg.2002.1132
17. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *NeuroImage.* (2012) 62:782–90. doi: 10.1016/j.neuroimage.2011.09.015
18. Cox RW. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res Int J.* (1996) 29:162–73. doi: 10.1006/cbmr.1996.0014
19. Cox RW. Afni: what a long strange trip it's been. *NeuroImage.* (2012) 62:743–7. doi: 10.1016/j.neuroimage.2011.08.056
20. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, et al. Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme. *Magnet Resonan Med.* (2009) 62:1609–18. doi: 10.1002/mrm.22147
21. Vidhusha S, Anandhan K, editors. Analysis and evaluation of autistic brain Mr images using learning vector quantization and support vector machines. In: *International Conference on Industrial Instrumentation & Control.* (2015). doi: 10.1109/IIC.2015.7150873
22. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *Stud Media Commun.* (1973) 3:610–21. doi: 10.1109/TSMC.1973.4309314
23. Haralick RM. Statistical and structural approaches to texture. *Proc IEEE.* (2005) 67:786–804. doi: 10.1109/PROC.1979.11328
24. Dalal N, Triggs, B, editors. Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision & Pattern Recognition.* (2005). doi: 10.1109/CVPR.2005.177
25. Toews M, Wells W III, Collins DL, Arbel T. Feature-Based morphometry: discovering group-related anatomical patterns. *NeuroImage.* San Diego, CA (2010) 49:2318–27. doi: 10.1016/j.neuroimage.2009.10.032
26. Castellani U, Rossato E, Murino V, Bellani M, Rambaldelli G, Perlini C, et al. Classification of schizophrenia using feature-based morphometry. *J Neural Transm.* (2012) 119:395–404. doi: 10.1007/s00702-011-0693-7
27. Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, et al. Investigating the predictive value of whole-brain structural Mr scans in autism: a pattern classification approach. *NeuroImage.* (2010) 49:44–56. doi: 10.1016/j.neuroimage.2009.08.024
28. Lao Z, Shen D, Xue Z, Karacali B, Resnick SM, Davatzikos C. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage.* (2004) 21:46–57. doi: 10.1016/j.neuroimage.2003.09.027
29. Etzel JA, Valchev N, Keysers C. The impact of certain methodological choices on multivariate analysis of fmri data with support vector machines. *NeuroImage.* (2011) 54:1159–67. doi: 10.1016/j.neuroimage.2010.08.050
30. Richiardi J, Eryilmaz H, Schwartz S, Vuilleumier P, Van De Ville D. Decoding brain states from Fmri connectivity graphs. *NeuroImage.* (2011) 56:616–26. doi: 10.1016/j.neuroimage.2010.05.081
31. Acharya UR, Ng EY, Tan JH, Sree SV. Thermography based breast cancer detection using texture features and support vector machine. *J Med Syst.* (2012) 36:1503–10. doi: 10.1007/s10916-010-9611-z
32. Klein A, Ghosh SS, Avants B, Yeo BT, Fischl B, Ardekani B, et al. Evaluation of volume-based and surface-based brain image registration methods. *NeuroImage.* (2010) 51:214–20. doi: 10.1016/j.neuroimage.2010.01.091
33. Yahaya K, Zawiah S, Zadry H. Medical image registration: Comparison and evaluation of nonlinear transformation algorithms. In: *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES).* (2010). doi: 10.1109/IECBES.2010.5742209
34. Robbins S, Evans AC, Collins DL, Whitesides S. Tuning and comparing spatial normalization methods. *Med Image Anal.* (2004) 8:311–23. doi: 10.1016/j.media.2004.06.009
35. Smith SM. Fast robust automated brain extraction. *Hum Brain Map.* (2002) 17:143–55. doi: 10.1002/hbm.10062
36. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
37. Cavedo E, Suppa P, Lista S, Lange C, Hampel H, editors. Impact of the Mri acquisition protocol on the hippocampus volume obtained by two methodologically distinct methods for fully automated volumetry: Spm12 and Fsl-first. *Alzheimers Dement*. (2015) 11:P689. doi: 10.1016/j.jalz.2015.06.1530
38. Zhu Z, Cao Y, Qin C, Rao Y, Ni D, Wang Y. Unsupervised 3d end-to-end deformable network for brain Mri registration. *Ann Int Conf IEEE Eng Med Biol Soc.* (2020) 2020:1355–9. doi: 10.1109/EMBC44109.2020.9176475
39. Zhang F, Wells WM, O'Donnell LJ. Deep diffusion mri registration (Ddmreg): a deep learning method for diffusion Mri registration. *IEEE Trans Med Imaging.* (2021). doi: 10.1109/TMI.2021.3139507

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Yuan, Tang and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 07 June 2022
doi: 10.3389/fpubh.2022.769692

A Neural Network and Optimization Based Lung Cancer Detection System in CT Images

Chapala Venkatesh[1], *Kadiyala Ramana*[2*], *Siva Yamini Lakkisetty*[1], *Shahab S. Band*[3*], *Shweta Agarwal*[4] *and Amir Mosavi*[5,6,7*]

[1] Department of ECE, Annamacharya Institute of Technology and Sciences, Rajampet, India, [2] Department of IT, Chaitanya Bharathi Institute of Technology, Hyderabad, India, [3] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliou, Taiwan, [4] SAGE University, Indore, India, [5] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [6] Faculty of Civil Engineering, TU-Dresden, Dresden, Germany, [7] Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
G Surya Narayana,
Vardhaman College of Engineering, India
Chennareddy Vijay Simha Reddy,
Middlesex University, United Kingdom
Venkata Subbaraju Dommaraju,
University of the Cumberlands, United States

***Correspondence:**
Shahab S. Band
shamshirbands@yuntech.edu.tw
Amir Mosavi
amir.mosavi@kvk.uni-obuda.hu
Kadiyala Ramana
ramana.it01@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 02 September 2021
Accepted: 20 January 2022
Published: 07 June 2022

Citation:
Venkatesh C, Ramana K, Lakkisetty SY, Band SS, Agarwal S and Mosavi A (2022) A Neural Network and Optimization Based Lung Cancer Detection System in CT Images.
Front. Public Health 10:769692.
doi: 10.3389/fpubh.2022.769692

One of the most common causes of death from cancer for both women and men is lung cancer. Lung nodules are critical for the screening of cancer and early recognition permits treatment and enhances the rate of rehabilitation in patients. Although a lot of work is being done in this area, an increase in accuracy is still required to swell patient persistence rate. However, traditional systems do not segment cancer cells of different forms accurately and no system attained greater reliability. An effective screening procedure is proposed in this work to not only identify lung cancer lesions rapidly but to increase accuracy. In this procedure, Otsu thresholding segmentation is utilized to accomplish perfect isolation of the selected area, and the cuckoo search algorithm is utilized to define the best characteristics for partitioning cancer nodules. By using a local binary pattern, the relevant features of the lesion are retrieved. The CNN classifier is designed to spot whether a lung lesion is malicious or non-malicious based on the retrieved features. The proposed framework achieves an accuracy of 96.97% percent. The recommended study reveals that accuracy is improved, and the results are compiled using Particle swarm optimization and genetic algorithms.

Keywords: cancer, lung cancer, machine learning, artificial intelligence, deep learning, cancer detection

INTRODUCTION

The most well-known reason for death because of malignant growth is lung cancer. The second most habitually analyzed type of malignancy is lung cancer. Pneumonic nodules are apparent in the lung to evaluate metastases from different malignancies (1, 2). Computed tomography (CT) is the most significant image mode for assessing progress/crumbling and for observation and decision-making malignant lung growths. As a result of the precocious presentation of lung malignancy by CT, doctors can be suggested more productive treatments (3, 4). Guess and recuperating components for scattered sickness with precise malignancy stages are required for orderly and consoling treatment (5). The early conclusion of the period of lung malignancy is firmly connected to the patient's continuance rate (6). In clinical terms, the disease is known to be strange hyperplasia and significantly beyond what 200 sorts can influence the individuals (7). According to the ACS (American Cancer Society), lung malignancy is the main cause of death in both men and women in the United States. about a total of 2,28,820 new lung malignancy cases were estimated, with 1,35,720 deaths (8). It causes a larger number of deaths than other malignant tumors. Early recognition of

tumorous lung nodules is the key factor for patient survival rate. When contrasted with chest X-ray imaging, CT perceives the tumorous nodules consistently at an underlying stage (9). Practically all radiologists use CT by exploring multiple pictures from a solitary patient. Thus, the exhaustion of the radiologists can prompt wrong analysis. Hence, the exact physical valuation measure is tedious and colossally inconsistent (10). A precise segment is noteworthy for the right valuation of nodule improvement and for the arrangement of malignant nodules (disease cells) from benign ones (non-disease cells). The reason for this work is to exactly recognize the nodules over the CT lung pictures.

The proposed instructional method pulls back, utlilizing a middle method to reduce confusion based on the CT image. Second, a crossover division approach is utilized to isolate the lung zone from its environmental factors. The proposed division strategy utilizes the Otsu thresholding to eliminate superfluous groups, consequently isolating the specific lung locales, and nodules of interest can be decisively characterized by cuckoo inquiry advancement. Third, the assortment of surface highlights for the particular nodule is undisturbed by parallel neighborhood examples at the feature stage. Finally, the highlights of the sectioned lung nodules are prepared by the CNN classifier to distinguish the lesions as malignant or non-malevolent.

RELATED WORK

In 2019, Ananya et al. (11) developed a multi-approach system for lung cancer categorization using genetics. They assessed false negatives and true positives for classification accuracy in this study, but not detected accuracy.

In 2019, Venkatesh et al. (12) developed an innovative approach to detect lesions based on a GA and LBP. This process achieves an accuracy of 90%.

In 2019. Preeti et al. (13) introduced a lung cancer detection framework based on the fuzzy c-mean clustering and SVM classifier techniques

In 2019, Senthil Kumar et al. (14) introduced an approach for detecting lung lesions using GCPSO. In this work, multiple optimization techniques are used to classify cancer in CT images. The process obtained a precision of 95%.

In 2018, Perumal et al. (15) proposed an ABC algorithm for malignancy recognition and classification. This guidance attained a 92% accuracy.

In 2017 Ammar et al. (16) established an early diagnostic architecture for genetically altered tumor detection. In this study, the authors achieved an accuracy rate of 84%.

In 2017 Kamil et al. (17) introduced a DWT-based lung lesion detection system. In this method by using subtraction and erosion techniques images are analyzed to remove the cancer region. This approach yielded an accuracy of 89%.

In 2016 Mukesh et al. (18) introduced a DWT-based method for assessing a high volume of tissues in chest X-ray images. Using this method, the authors were able to achieve an accuracy of 86%.

In 2014 Santos et al. (19) described an area development and Hessian matrix to identify minor respiratory lesions. The presented approach achieves a classification accuracy of 88.4%.

In 2014, Jinsa and Gunavathri (20) reported an ANN-based lesion categorization technique. They were able to classify with an accuracy of 93.3%.

The principal gap has indeed been extended due to the lack of research publications that demonstrate computations. Because of poor directionality, slower processing, greater calculation time, and complex computations, the methods suggested by the authors mentioned above are less effective in all cases. As a result, an interactive technique for identifying lung cancer in CT images is suggested in this article, which uses the otsu threshold-based Cuckoo search algorithm, Local Binary Pattern for image retrieval, and CNN for classification to conquer all of the shortfalls of the existing methods. By selecting the most cost-effective strategy, optimization algorithms tend to produce a solution for image processing processes.

MOTIVATION AND CONTRIBUTION

Lung cancer is confirmed by physicians after a thorough examination of CT scans, which requires a lot of time and is not always accurate. To create imagery as precise, operational, and efficient as possible, state-of-the-art optimization techniques and image processing approaches were required. The proposed technology will aid doctors in accurately identifying lung nodules at an early stage, as well as studying the internal anatomy. As a part of the contribution, some glitches related to lung cancer detection are discussed here. The region of interest is retrieved using Otsu thresholding and cuckoo search optimization, which is a novel approach to segmentation. This proposed partitioning approach requires only a few parameters to precisely separate nodules of varied sizes and shapes.

Proposed Methodology

Figure 1 depicts the prospective lung malignancy diagnostic procedure, which comprises five phases: (1) contrast enhancement and Noise reduction through pre-processing, (2) Otsu thresholding based cuckoo search algorithm to segment the lesion from its backgrounds, (3) retrieval of regions of concern, (4) retrieval of descriptors from segmented lung lesions, and, in the last phase, (5) SVM has been used to assess if the lesion was abnormal or normal. The next sections provide detailed descriptions of the above-mentioned phases.

Image Acquisition

It is the basic step before proceeding with other critical steps. It is a method for processing a digital image from a database (21). Numerous sorts of scanners, such as X-Ray, MRI, and CT, are used to obtain the images. The CT image was captured using a CT scanner. It is a type of scanning that creates cross-section scans for each pixel (22).

Pre-processing

The equations should be inserted in editable format from the equation editor. It is then procedure to improve image details. The basic idea is to suppress noise, which corrects undesired distortions and enhances the associated attributes of the image for subsequent processing (23). Because all techniques are sensitive to noise, efficient images pre-processed allow for better

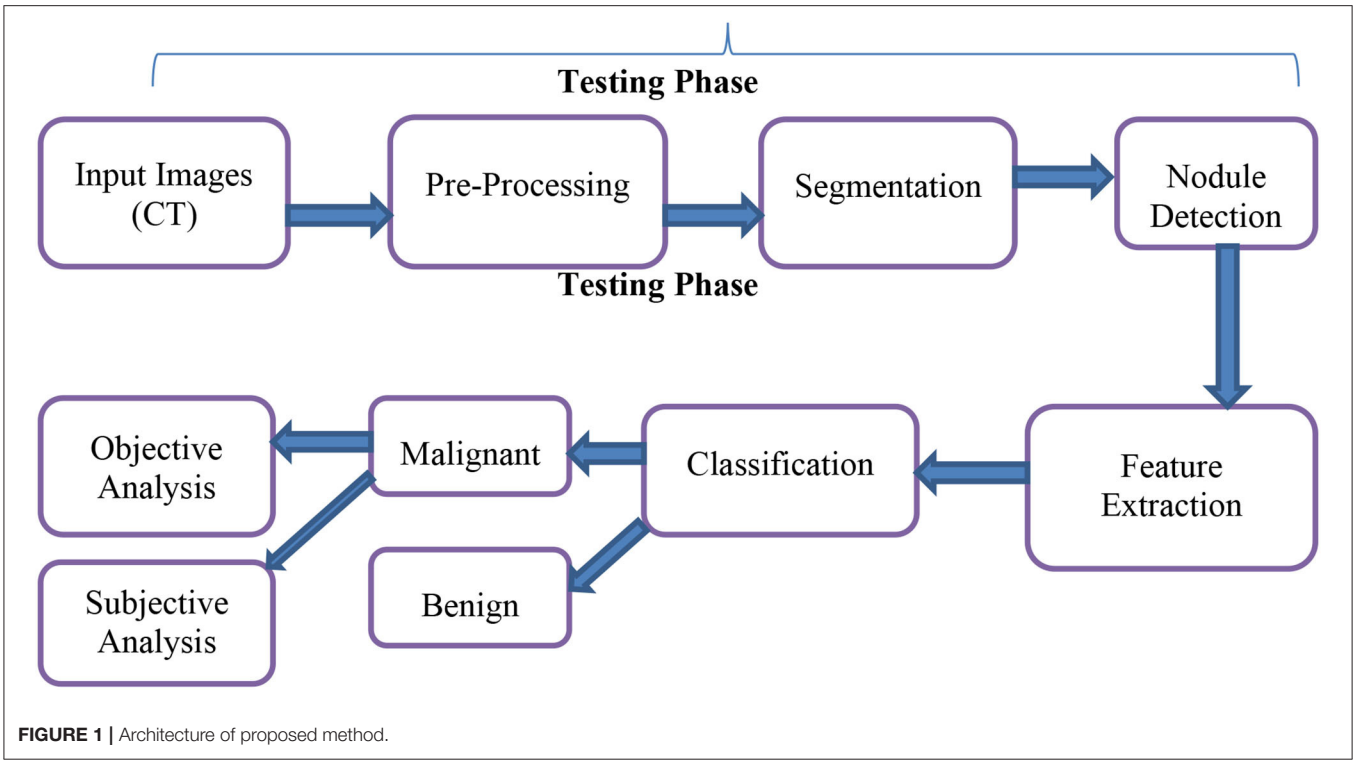


FIGURE 1 | Architecture of proposed method.

segmentation and, as a result, better classification. The size of the pixel area could be used to classify pre-processing procedures. Image enhancement employs these techniques. Enhancement operations operate on the image pixels of the neighborhood and the corresponding values of the neighborhood. Contribute quality to the images by decreasing noise and distortion (24).

$$\mathbf{f}(x) = \mathbf{a}_0 + \sum_{n=1}^{\infty} \left(\mathbf{a}_n \cos \frac{n\pi x}{\mathbf{L}} + \mathbf{b}_n \sin \frac{n\pi \mathbf{x}}{\mathbf{L}} \right)$$

Median Filtering

Salt and pepper noise can be found on CT scans. The finest features are obscured by these impacts. By keeping the frontier of the image as fine as possible, this filtering lowers salt and pepper noise (12). This filter gathers information from a sample within a non-averaged window (25). The edges of the filter are better managed than those of other linear filters. The following equations are used to get the median value.

$$\mathbf{M}(g) = \sum_{k=1}^{n} |x_k - \mathbf{g}| . \quad (1)$$

$$\mathbf{g} = \mathbf{median}\{x_1, x_2, \ldots\ldots\ldots x_n\} . \quad (2)$$

Otsu Thresholding Segmentation With Optimization

Otsu Thresholding

The goal of this strategy is to scour specified like-classes of pixels in a picture for the closeness of neighboring pixels in order to generate a concentrated image object. Separating the background and sub-regions in medical imaging is tough (26). The Otsu segmentation algorithm works better to "recognize" or "smear" the context contents of the front objects. It is an adaptive threshold binarization procedure proposed by OTSU in 1979. This procedure uses the highest within-class variance between the context and the target based on the rule of threshold assortment (27). It segments the image into the forefront and the contextual based on the characteristics of gray level values. If the finest threshold is attained, the gap between the two regions is the highest. The Otsu algorithm, in general, utilizes the greatest within-class variance. The larger the variance value, the wider the difference between the two areas, since variance is a useful determinant of uniform gray distribution. If some areas are wrongly segmented into contextual or if some contextual is segmented into areas, then the gap is too small between the two areas. As a result, if the variance among groups is higher, the likelihood of incorrect classification is lowered, resulting in cohesive segmentation.

The following is the main principle of OTSU-based threshold segmentation:

Let us call the gray values g and the number of pixels nx. Then

$$\mathbf{P} = \sum_{x=0}^{\mathrm{L}-1} n_x = n_0 + n_1 + n_2 + \ldots + n_{\mathrm{L}-1} \ldots \quad (3)$$

where g = 0, 1,..., L-1, and P indicate the number of pixels. Suppose C1 and C2 are the two kinds of pixels. C1 pixels have

a range of [0,x] while C2 pixels have a range of [x+ 1,L1].

$$\sigma_{Gv}^2 = \sum_{g}^{L-1} (g - m_{Gv})^2. \quad (4)$$

$$\sigma_{Bv}^2 = P_1(m_1 - m_{Gv})^2 + P_2(m_2 - m_{Gv})^2 \quad (5)$$

The below-mentioned calculations are used to compute the mean intensities.

$$m_1 = \frac{1}{P} \sum_{g}^{x} g.P_g \quad (6)$$

$$m_2 = \frac{1}{P} \sum_{g=x+1}^{L-1} g.P_g \quad (7)$$

$$m_{Gv} = \sum_{g=0}^{L-1} g.P_g \quad (8)$$

where m 1 and m 2 are the C1 and C2 pixel average intensities, and m Gv is the global mean intensity. Lastly, the ratio τ , which is provided below, is used to determine the ideal threshold.

$$\tau = \frac{\sigma_{Bv}^2}{\sigma_{Gv}^2} \quad (9)$$

Cuckoo Search Optimization

The cuckoo generation function is projected using this approach, which reduces the implications. A large number of nests are accessible during the search procedure. The location of the cuckoo egg has been discovered as a novel solution (28). The steps in the search procedure are as follows. A cuckoo bird places one egg at a time in a randomly chosen nest. The parasite nests were static, and the number of eggs in the nests would increase until they reach their highest level. When the cuckoo's egg is spotted, the host bird seems to have the choice of chucking the egg away or scrapping the nest and forming a new one.

The Levy flight theory has improved the CS algorithm (29). This CS technique is used to calculate the appropriate threshold for eliminating the lung nodule.

The following analogy is incorporated into the proposed technique for optimum selection (**Algorithm 1**):

The new solution, which depicts the subcategory of thresholds, is depicted by the egg of a cuckoo. This is also utilized for segmentation of the lung nodule. The grade of eggs for each host nest is either 0 or 1, which mimics the segmentation procedure's threshold partition. Pa is the probability that a cuckoo's egg will be discovered by the host bird. Pa has a predefined threshold. It demonstrates the principle of removing the least relevant threshold subgroups and, as a result, removing these threshold values from further analysis.

Feature Extraction

The LBP operator was established to determine texture in the first place (30). By thresholding an image with the central pixel value and taking the result as a binary quantity, the operator applies a mark to each pixel (31). The picture of the Lung CT could be considered as a micro-pattern structure that the LBP operator can well portray. The steps for extracting the characteristics are outlined below.

Algorithm 1 | Cuckoo search algorithm.

Step1: Initialization parameters: n, Pa, & M where n=number of host nests; pa : probability of discovery of alien, M: maximum number of iterations

Step2: Generate initial n host, n_i^t

Step3: Evaluate f(n_i^t)

Step4: Generate a new solution $n_i^{t+1} = n_i^t + \alpha \oplus Lev'y(\gamma)$

Where the symbol $\oplus$ is entry-wise multiplication,
$\alpha > 0$ indicates the step size, Levy(γ)= $g^{-\gamma}(1 < \gamma \leq 3)$

Step5: Evaluate $f(n_i^{t+1})$

Step6: Choose a nest $n_j$randomly

Step7: If $(n_j^t) > (n_j^{t+1})$ then Replace n_j^t *with* n_j^{t+1}

Step8: Confiscate a worse nest with Pa

Step9: Construct new nest using Levy flights

Step10: Retain the best solutions

- Divide the window looking at into cells.
- In a cell, collate each pixel with its neighbors.
- If the value of the center pixel is larger than the value of the neighbor pixel, assign "1"; otherwise, assign "0".
- A binary number is created by comparing all of the pixels.
- Lastly, over the cell, compute the histogram. The LBP value can be calculated using the expression $LBP_{X,Y}^U$ where U represents uniform pattern and X,Y indicate neighborhood.

$$LBP_{X,Y} = \sum_{0}^{m-1} s(P_s - P_c)2^s \quad (10)$$

$$s(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (11)$$

P_c is the gray value of the center pixel, P s is the intensity value of the location pixels (m = 0, 1,.....m-1), and m is the image element well within range R where R is higher than zero (R > 0), creating a regionally oriented neighborhood set. After identifying each pixel in a picture, a histogram is created to define the texture image (32, 33).

CLASSIFICATION

CNN belongs to DNN group which is comprised of numerous hidden layers, like RELU, fully linked, pooling and convolution layer, etc. CNN securities weights in the coevolutionary layer, which lowers the network latency, and enhances network performance (34). CNN's prominent features are prevalent weights, local networking, and neuronal 3D sizes. A feature map is created with a kernel by a convolution layer of diverse sub-regions of the input image (35). Then, a nonlinear function is added to the RELU layer to progress the convergence possessions when the error is small. The architecture of CNN is as shown in **Figure 2**.

In relation to complex layers, CNNs quite often incorporate pooling layers (36). They are principally used only for lessening the dimensions of the tensor and speeding up estimations (37). All such layers are simple. So, the image

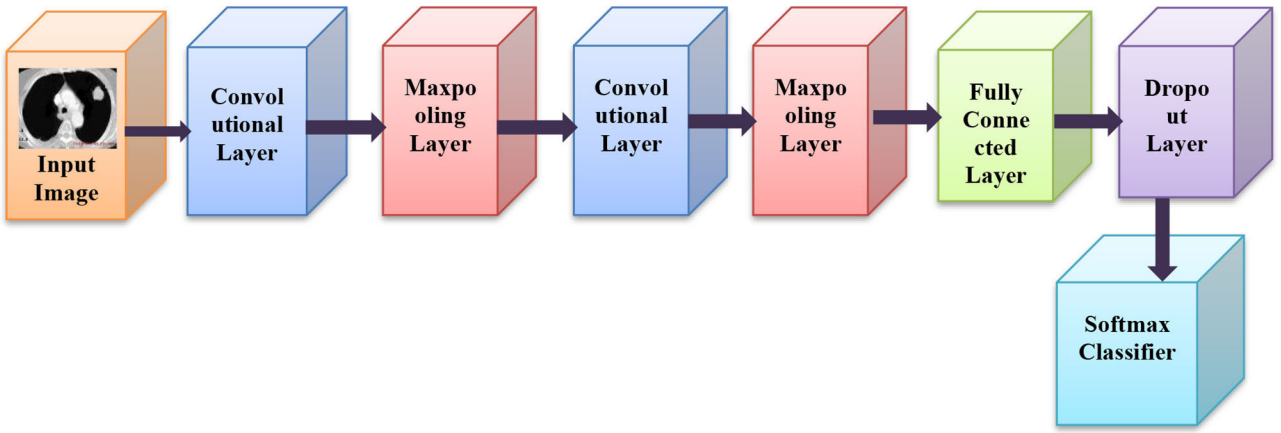


FIGURE 2 | Architecture of CNN.

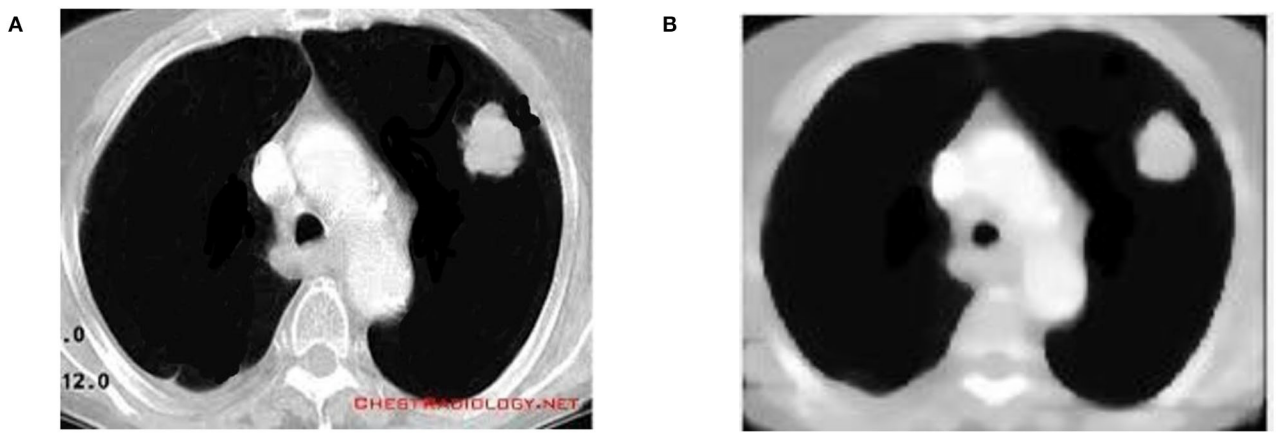


FIGURE 3 | (A) Input CT image. **(B)** Filtered output.

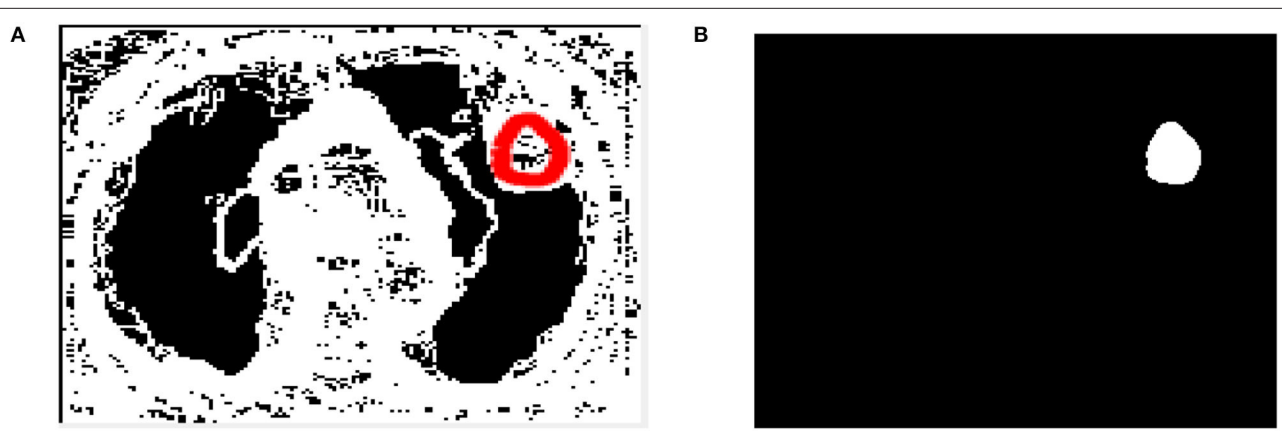


FIGURE 4 | (A) Extracted output. **(B)** Segmented output.

is to be split up into smaller portions in the pooling layer and for each portion, the maximum value is selected and then accomplished in some process for each portion (38). After being portioned, it is placed in the output in the respective position. RELU is a rectified linear unit, as well as a form of hidden layers. The activation function is

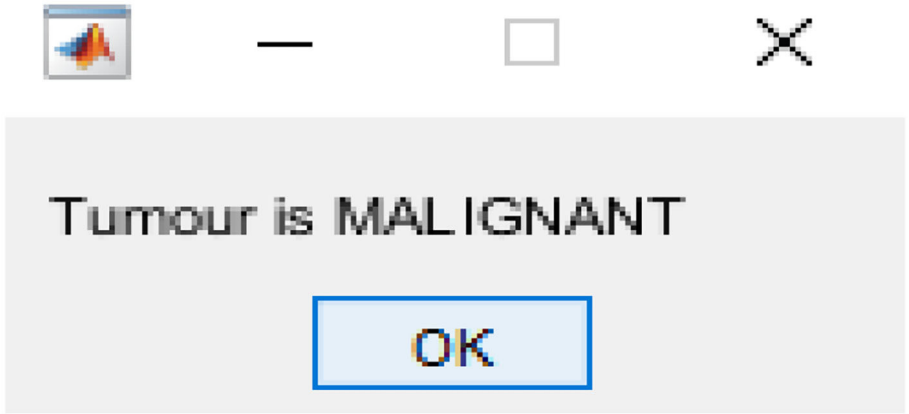


FIGURE 5 | Classification output.

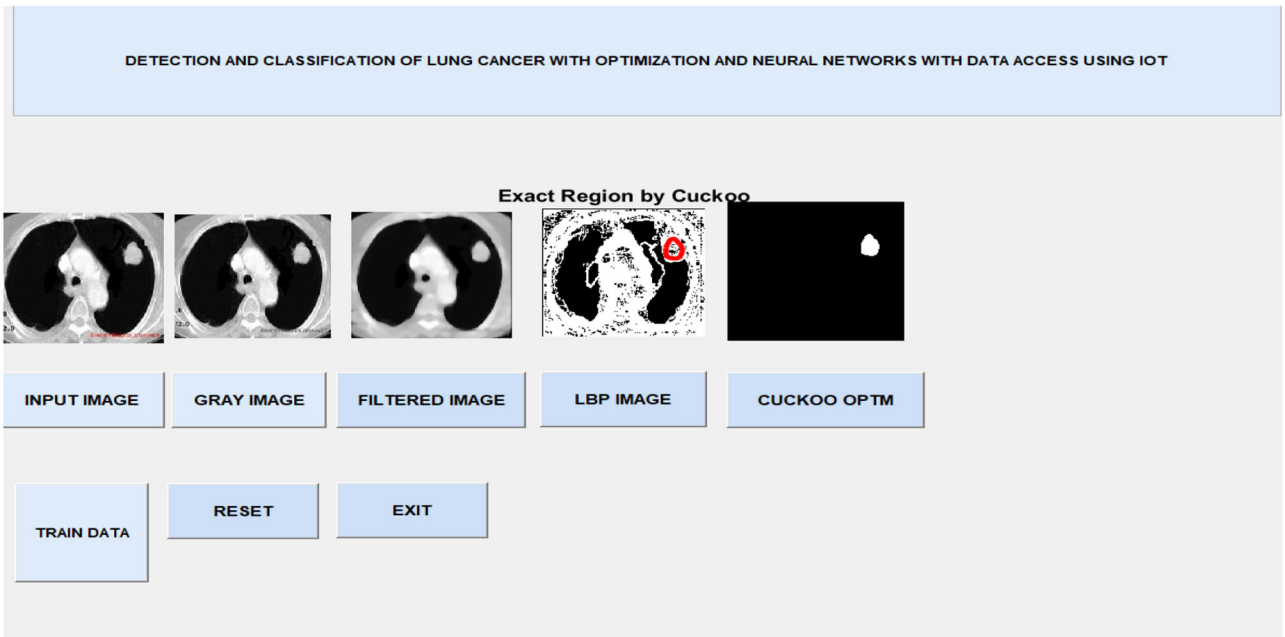


FIGURE 6 | GUI output.

most popularly used in neural networks, predominantly in CNNs (34).

SIMULATION RESULTS

Lung cancer is diagnosed in CT medical images using the novel cuckoo search algorithm and attributes are determined in this study. Lung cancer CT images were collected from a private hospital (Satyam diagnostic center, Anantapur). The adaptive threshold issue in this study is referred to as an optimization problem and it can be resolved using the CSA approach. In this study, the outcomes of the suggested method were compared to those of the PSO and GA algorithms. This work has been carried out by MATLAB software. When compared to open-source tools, MATLAB has a great affinity with deep learning techniques as well as hardware tools. Also, open-source tools have a hard time bringing all of the libraries together in one spot.

This study relies on my prior work (39), in which the outcomes were produced using a Genetic algorithm and particle swarm optimization approaches in addition to LBP and CNN. Cuckoo search optimization is used in this work, together with CNN and LBP, to enhance the accuracy. The input and median filter output of CT lung cancer pictures are depicted in **Figure 3**. Low-frequency noise

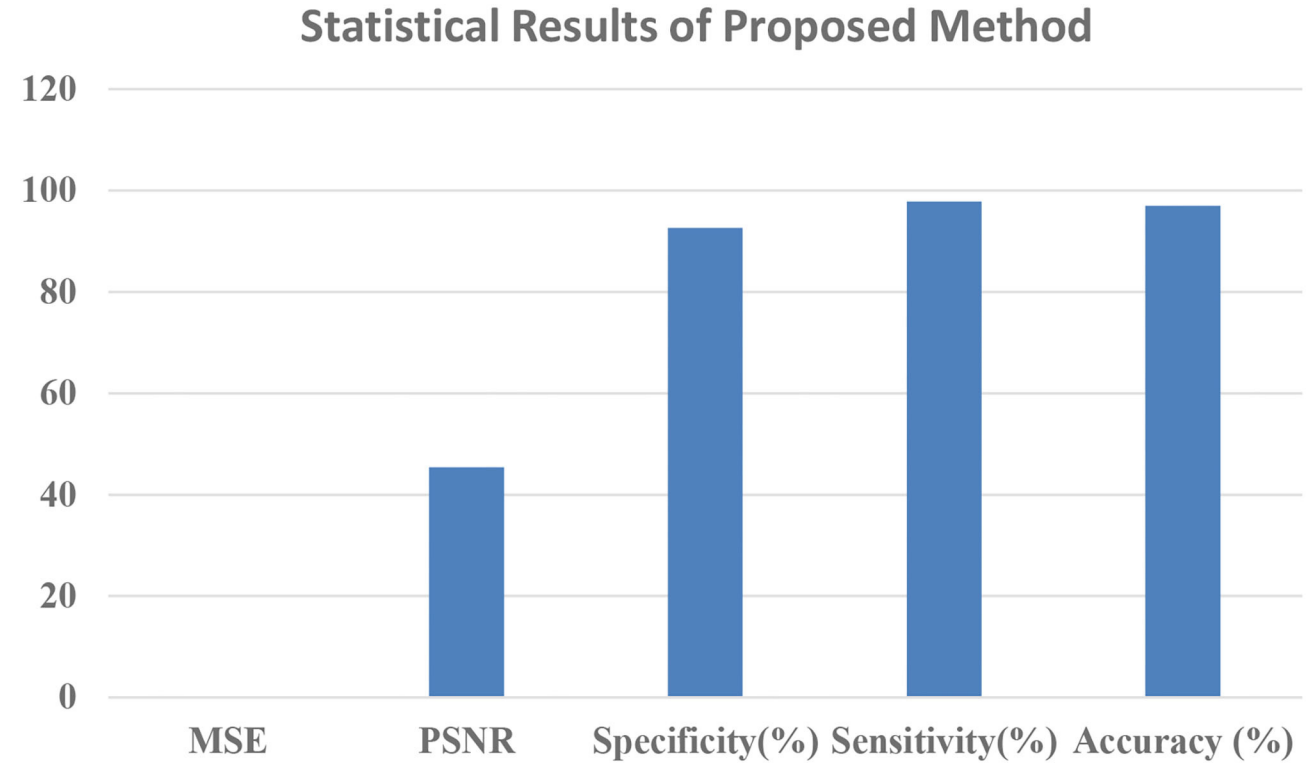


FIGURE 7 | Statistical results graphical representation.

and distortion are common in CT scan images. To reduce noise and distortion, the input image is processed with a median filter.

To identify lung lesions in the CT image, it is split into multiple clusters and then optimized using the Otsu thresholding approach. The CT scan is first split using simple Otsu thresholding, which improves the segmented classes' variation, or "all class variance." The result of the thresholding technique can be enhanced by processing it with cuckoo search optimization. Following partitioning, the image is subjected to LBP feature extraction, which extracts the textural features before extracting the detected output (see **Figure 4**).

After image retrieval, the image is given to CNN classification, which assesses the image as normal or abnormal by showing a message such as "Tumor is MALIGNANT" or "Tumor is BENIGN," as illustrated in **Figure 5**. The system's general function is created in a GUI, as seen in **Figure 6**. The statistical results like performance metrics obtained by 200 iterations for the input image are shown in graphical representation in **Figure 7**.

Table 1 showed that the suggested approach yields higher accuracy of 97%, the sensitivity of 97.8%, specificity of 92.6%, PSNR of 45.38%, and low MSE of 0.013 than conventional systems. These optimum results are obtained for 200 iterations (Our earlier proposed systems).

TABLE 1 | Attributed obtained from the proposed method.

Parameters	Proposed Method (CSO+CNN+LBP)
MSE	0.013
PSNR (%)	45.38
Specificity (%)	92.672
Sensitivity (%)	97.806
Accuracy (%)	96.979

TABLE 2 | Comparative Results with proposed method.

Parameters	Proposed Method (CSO+CNN+LBP)	(PSO+SVM+LBP)	(GA+SVM +LBP)
MSE	0.013	0.0301	0.0651
PSNR	45.38	33.2788	27.5311
Specificity (%)	92.672	60.0000	90.4950
Sensitivity (%)	97.806	96.5783	83.7143
Accuracy (%)	96.979	96.9391	90.4937

Table 2 shows that the suggested approach has higher accuracy (97%) than conventional systems (Our earlier proposed systems). The Comparative Results Graphical Representation is shown in **Figure 8**.

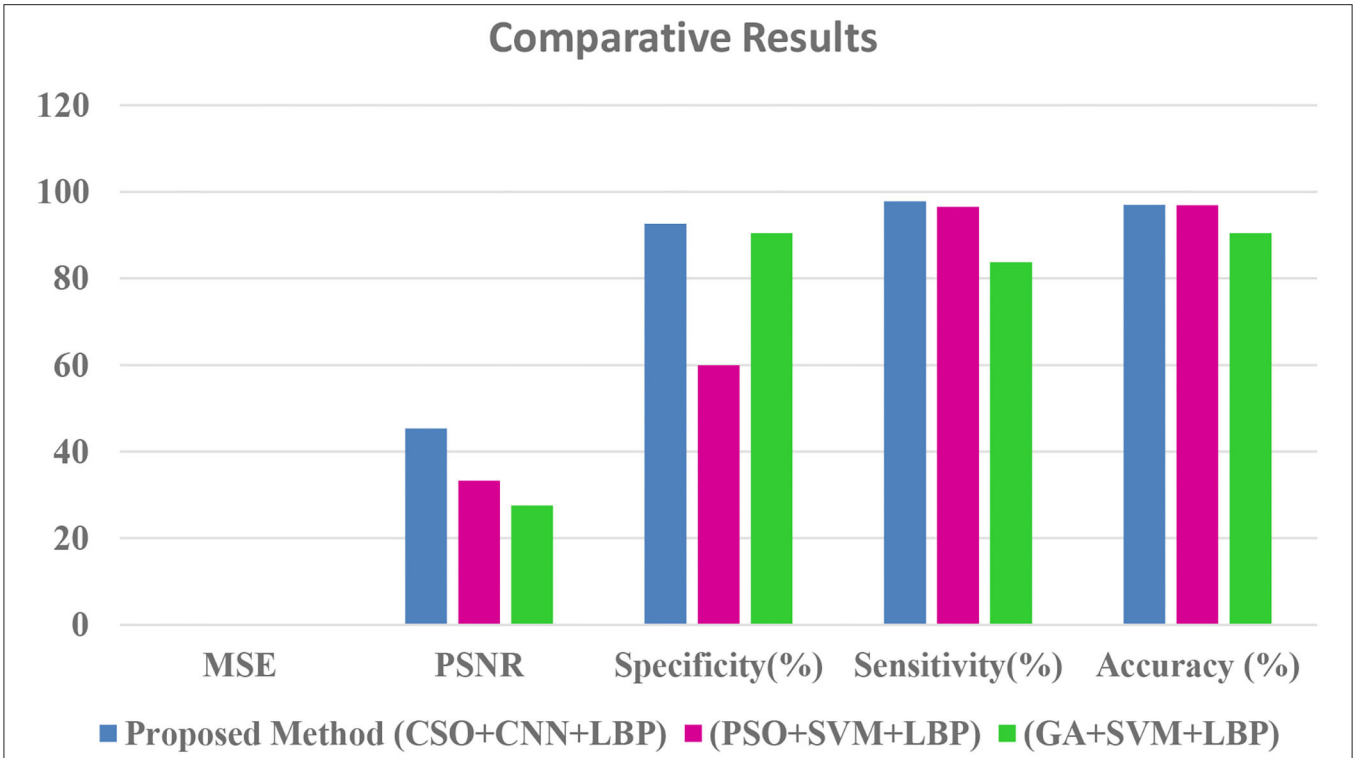


FIGURE 8 | Comparative results graphical representation.

CONCLUSION

In this article, a strong approach for recognizing lung cancer in CT images is developed. For exact cancer diagnosis in CT lung images, the Otsu thresholding-based cuckoo search optimization and CNN classifier approach were used. Based on the simulation findings, it is observed that the suggested method reliably segments CT images and detects lesions of various forms and sizes. Subsequently, the proposed approach comprises successive stages that continuously yield the last detection result. The techniques that are utilized in different stages are basic and simple to actualize.

Based on the simulation findings, the accuracy of the proposed framework' is calculated to be 96.97%, which is greater than any other demonstrative framework found in the literature. As for future work, a powerful strategy could be created by supplanting the CNN Classifier with a profound deep learning method and CAD tools. One may improve the structure for images of lung cancer in different modalities.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CV: conceptualization, methodology, software, and visualization. KR: data curation, writing—original draft, data analysis, and investigation. SL: software, validation, and editing. SB: software, validation, and editing. AM: supervision, writing—review & editing. SA: software, validation, and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors express sincere thanks to Dr. M. Vijay Kumar, Radiologist from Star Diagnostics Hospital at Ananthapuramu, for providing the lung CT images to carry-out this research work.

REFERENCES

1. Manickavasagam R, Selvan S. Automatic detection and classification of lung nodules in CT image using optimized neuro fuzzy classifier with cuckoo search algorithm. *J Med Syst.* (2019) 43:1–9. doi: 10.1007/s10916-019-1177-9
2. Brown MS, McNitt-Gray MF. Patient-specific models for lung nodule detection and surveillance in CT images. *Proc IEEE Trans Med Imag.* (2001) 20:1242–50. doi: 10.1109/42.974919
3. Cosman PC, Tseng C, Gray RM. Tree-structured vector quantization of CT chest scans: image quality and diagnostic accuracy. *Proc IEEE Trans Med Imag.* (1993) 12:727–39.
4. Dewes P, Frellesen C, Al-Butmeh F, Albrecht MH, Scholtz J-E, Metzger SC, et al. Comparative evaluation of non-contrast CAIPIRINHA-VIBE 3T-MRI and multidetector CT for detection of pulmonary nodules: in vivo evaluation of diagnostic accuracy and image quality. *Eur J Radiol.* (2016) 85:193–8. doi: 10.1016/j.ejrad.2015.11.020

5. Georg Homann MD, Mustaf DF. Improved detection of bone metastases from lung cancer in the thoracic cage using 5- and 1-mm axial images versus a new CT software generating rib unfolding images. *J Acad Radiol*. (2015) 22:505–12. doi: 10.1016/j.acra.2014.12.005
6. De Nunzio G, Massafra A. Approaches to juxta-pleural nodule detection in CT images within the MAGIC-5 collaboration. *J Nucl Instrum Methods Phys Res*. (2011) 648:103–6. doi: 10.1016/j.nima.2010.12.082
7. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Lwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. *IET Image Process*. (2021) 16:647–58. doi: 10.1049/ipr2.12222
8. American Cancer Society. *Key Statistics for Lung Cancer*. Available online at: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html (accessed January 5, 2022).
9. Prokop M, Galanshi M. *Spiral and Multislice Computed Tomography of the Body*. Stuttgart: Thime Medical Publishers (2003)
10. Abbas S, Jalil Z, Javed AR, Batool I, Khan MZ, Noorwali A, et al. BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm. *Peer J Comput Sci*. (2021) 7:e390. doi: 10.7717/peerj-cs.390
11. Ananya C, Rajamenakshi, Subramanian R, Gaur S. A novel approach for tumor segmentation for lung cancer using multi-objective genetic algorithm and connected component analysis. In: *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology, Advances in Intelligent Systems and Computing Springer Nature*. Singapore (2019). pp. 367–76.
12. Venkatesh C, Polaiah B. A novel approach for lung lesion segmentation using optimization technique. *Helix Scientific Explor*. (2019) 94:4832–7. doi: 10.29042/2019-4832-4837
13. Preeti J, Bajaj SB, Aman J. "Segmentation and Detection of Lung Cancer Using Image Processing and Clustering Techniques" Springer Nature Singapore, Progress in Advanced Computing and Intelligent Engineering. *Adv Intell Syst Comput*. (2019) 1:3–23. doi: 10.1007/978-981-13-1708-8_2
14. Senthil Kumar K, Venkatalakshmi K, Karthikeyan K. Lung cancer detection using image segmentation by means of various evolutionary algorithms. *Hindawi Comput Math Methods Med*. (2019) 2019:1–6. doi: 10.1155/2019/4909846
15. Perumal S, Velmurugan T. Lung cancer detection and classification on CT CT scan images using enhanced artificial bee colony optimization. *Int J Eng Technol*. (2018) 7:74–9. doi: 10.14419/ijet.v7i2.26.12538
16. Ammar O, Ibrahim Al A, Abrahim B. Novel genetic algorithm for early prediction and detection of lung cancer. *J Cancer Treat Res*. (2017) 5:15–8. doi: 10.11648/j.jctr.20170502.13
17. Kamil D, Ali H, Yoney KE. Lung lesion segmentation using gaussian filter and discrete wavelet transform. *ITM Web Conf*. (2017) 11:1–10. doi: 10.1051/itmconf/20171101018
18. Mukesh CA Dr, Bhumika G. Detect mass tissue in lung images using discrete wavelet transformation. In*: Proceedings of the IEEE International Conference on Information Processing (IICIP)* (New Delhi), (2016):1–10.
19. Santos AM, Ode A, Filho C, Silva AC, Nunes RA. Automatic detection of small lung nodules in 3D CT data using gaussian mixture models, Tsallis entropy and SVM. *Eng Appl Artif Intell*. (2014) 36:27–39. doi: 10.1016/j.engappai.2014.07.007
20. Jinsa K, Gunavathi K. Lung cancer classification using neural networks for CT images. *Comput Methods Programs Biomed*. (2014) 113:202–209. doi: 10.1016/j.cmpb.2013.10.011
21. Anita P, Raja SK, Gandharba S. Digital image steganography using LSB substitution, PVD, and EMD. *Hindawi Math Probl Eng*. (2018) 2018:1–11. doi: 10.1155/2018/1804953
22. Venkatesh C, Polaiah B. An investigation of diverse optimization techniques on medical imagery for detection of perilous diseases. *Front J Soc Technol Environ Sci*. (2017) 62:49—255. doi: 10.1108/IJPCC-10-2020-0160
23. Thippa Reddy G, Khare N, Bhattacharya S, Singh S, Maddikunta PKR, In-Ho Ra, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*. (2020) 9:274. doi: 10.3390/electronics9020274
24. Madhav BTP, Pardhasaradhi P,.Manepalli RKNR, Kishore PVV, Pisipat VGKM. Image enhancement using virtual contrast imagefusion on Fe3O4 and ZnO nanodispersed decyloxybenzoic acid. *Liq Cryst*. (2018) 42:1329–36. doi: 10.1080/02678292.2015.1050704
25. Prasanna MN, Polaiah B. Optimization of rotary kiln in cement industry using conventional control systems. *Helix*. (2019) 9:4843–49. doi: 10.29042/2019-4843-4849
26. Venkatesh C, Bhagyalakshmi K, Sivayamini L. Detection of diverse tumefactions in medial images by various cumulation methods. *Int Res J Eng Technol (IRJET)*. (2017) 4:1195–200.
27. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Sys Man Cyber*. (1979) 9:62–6.
28. Xin-She Y. *Cuckoo Search and Firefly Algorithm Theory and Applications*. Springer (2014).
29. Yang XS, Deb S. Cuckoo search via levy flights. In: *Proceedings of World Congress on Nature and Biologically Inspired Computing (NaBIC 2009)* (Coimbatore), (2009):210–14.
30. Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognit*. (1996) 29:51–9. doi: 10.1016/0031-3203(95)00067-4
31. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
32. Aparna P, Polurie V, Vijay K. An efficient medical image watermarking technique in e-healthcare application using hybridization of compression and cryptography algorithm. *J Intell Syst*. (2017) 27:115–33. doi: 10.1515/jisys-2017-0266
33. Malakonda Reddy B, Rahman ZUr Md. SAR electromagnetic image conditioning using a new adaptive particle swarm optimization. *ACES J*. (2018) 33:1439–46.
34. Kumar EK, Kishore PVV, Sastry ASCS, Teja Kiran Kumar M, Anil Kumar D. Training CNNs for 3-D sign language recognition with color texture coded joint angular displacement maps. *IEEE Signal Process Lett*. (2018) 25:645–9. doi: 10.1109/LSP.2018.2817179
35. Bhubaneswar P, Brintha Therese A. Detection of cancer in lung with K-NN classification using genetic algorithm. *Proc Mater Sci*. (2015) 10:433–440. doi: 10.1016/j.mspro.2015.06.077
36. Othayoth R, Mathi P, Bheemanapally K, Kakarla L, Botlagunta, M. Characterization of vitamin-cisplatin-loaded chitosan nano-particles for chemoprevention and cancer fatigue. *J Microencapsul*. (2015) 32:578–88. doi: 10.3109/02652048.2015.1065921
37. Svitlana K, Chauhdar SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-Technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access*. (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
38. Ratna Bhargavi V, Rajesh V. Computer aided bright lesion classification in fundus image based on feature extraction. *Intern J Pattern Recognit Artif Intell*. (2018) 32:1850034. doi: 10.1142/S0218001418500349
39. Venkatesh C, Bojja P. Development of qualitative model for detection of lung cancer using optimization. *Int J Innovative Technol Explor Eng*. (2019) 8:3143–47. doi: 10.35940/ijitee.I8619.078919

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Venkatesh, Ramana, Lakkisetty, Band, Agarwal and Mosavi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 10 June 2022
doi: 10.3389/fpubh.2022.911377

Effects of Osimertinib Combined With Pulmonary Rehabilitation and Health Care Training on Pulmonary Function, Complications, and Quality of Life in Patients After Radical Resection of Lung Cancer

Haijiang Xu[1], *Ruixia Guo*[2] *and Yantao Yang*[1]*

[1] Department of Pharmacy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Department of Gynecology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Objective: To explore the effects of osimertinib combined with pulmonary rehabilitation and health care training on pulmonary function, complications, and the quality of life (QOL) in patients after radical resection of lung cancer.

Methods: The data of 120 patients with radical resection of lung cancer admitted to *The First Affiliated Hospital of Zhengzhou University* from February 2020 to February 2021 were retrospectively analyzed. According to the order of admission, they were equally divided into group p and group q. All patients were given pulmonary rehabilitation and health care training, and group p was treated with osimertinib, while group q received the treatment of pemetrexed combined with cisplatin. The pulmonary function, the incidence of complications, and QOL between the two groups were compared.

Results: Compared with group q, the pulmonary function was higher ($P < 0.001$), the incidence of complications was significantly lower ($P < 0.05$), and QOL scores were markedly higher in group p after treatment ($P < 0.001$).

Conclusion: The combination of osimertinib and pulmonary rehabilitation and health care training can improve the pulmonary function of patients with non-small cell lung cancer (NSCLC) with radical resection of lung cancer, and reduce their postoperative morbidity, thereby improving their QOL, which is conducive to reducing the patient's and society's medical burden.

Keywords: osimertinib, pulmonary rehabilitation and health care training, radical resection of lung cancer, pulmonary function, quality of life

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Santhosh Krishna B. V., New Horizon College of Engineering (NHCE), India
Navod Thilakarathne, University of Colombo, Sri Lanka

***Correspondence:**
Yantao Yang
fccyangyt@zzu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *02 April 2022*
Accepted: *22 April 2022*
Published: *10 June 2022*

Citation:
Xu H, Guo R and Yang Y (2022) Effects of Osimertinib Combined With Pulmonary Rehabilitation and Health Care Training on Pulmonary Function, Complications, and Quality of Life in Patients After Radical Resection of Lung Cancer. Front. Public Health 10:911377. doi: 10.3389/fpubh.2022.911377

INTRODUCTION

Lung cancer is a malignant tumor originating from the mucosa or gland of the trachea and bronchus. According to the differences in histopathological characteristics, lung cancer is generally divided into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) in the clinic. NSCLC accounts for approximately 85% of all patients with lung cancer, while SCLC accounts

for only 15% (1, 2). In recent years, with changes in the social environment and living habits, the clinical characteristics of lung cancer have seen a substantial transformation in China (3, 4), but the main treatment option of patients has not been affected. Surgery is still the main treatment for patients with early, medium, and some advanced NSCLC, and 90% of patients are supplemented with chemotherapy after surgery (5) to improve QOL. At present, chemotherapy has an exact effect on SCLC, and about 1% of patients can be cured by chemotherapy (6). However, the effects of chemotherapy in patients with NSCLC are unsatisfactory, and patients can only relieve the clinical symptoms through chemotherapy drugs, which cannot effectively improve pulmonary function (7). For patients with NSCLC with the EGFR gene mutation, the T790M mutation will lead to an increase in chemoresistance, and patients who are insensitive to the conventional pemetrexed often get worse after treatment (8). In recent years, the third generation of epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI) developed by AstraZeneca has been used in the treatment of the NSCLC with the EGFR gene mutation. The drug has a significant therapeutic effect on the central nervous system with mild toxicity and favorable safety. The postoperative rehabilitation and health care training based on medication can effectively improve the endurance and strength of patients' respiratory muscles, thereby reducing the bodily oxygen consumption, enhancing the cardiopulmonary and muscular function, and providing a benign physiological basis for medication (9) and improving the patients' QOL. There is no study on the effects of osimertinib combined with pulmonary rehabilitation and health care training on patients with NSCLC with thee radical resection of lung cancer by reviewing the previous literature. Based on this, 120 patients were selected to explore the actual effects of osimertinib combined with pulmonary rehabilitation and health care training. The reports are as follows.

MATERIALS AND METHODS

Research Design

This retrospective study was conducted under the guidance of the Declaration of Helsinki (2013) (10) to explore the effects of osimertinib combined with pulmonary rehabilitation and health care training on pulmonary function, complications, and QOL in patients after radical resection of lung cancer. The blind level of this study was double-blind. The study subjects and researchers did not know the grouping of this experiment, and the research designers were responsible for arranging and controlling the experiment. This study was approved by the ethics committee of The First Affiliated Hospital of Zhengzhou University.

General Information

The data of patients with radical resection of lung cancer admitted to *The First Affiliated Hospital of Zhengzhou University* from February 2020 to February 2021 were retrospectively analyzed, and the patients were included according to the following criteria: (1) Patients with the pulmonary CT examination as punctiform or patchy shadows who were diagnosed with lung cancer by pathological examination, which met the diagnostic criteria in the Chinese guidelines on the diagnosis and treatment of primary lung cancer (2015 version) (11); (2) Patients who had the EGFR T790M mutation-positive non-small cell lung cancer (12); (3) Patients who had surgical indications for radical resection of lung cancer; (4) Patients who had no disease affecting the motor nerve, joint, and muscle; (5) Patients who were not malnourished; (6) Patients who had an expected survival time of more than 6 months; (7) The Karnofsky performance status score was more than 60 points in patients; (8) The age of the patients was more than 18 years old; and (9) Patients who had good tolerance and compliance. In the following circumstances, patients were not included in this study: (1) Patients with some factors such as hearing impairment, language disorders, unconsciousness or mental illness, and patients who could not communicate with others; (2) Patients with a grave organic disease or dysfunction in vital organs; (3) Patients with complications, such as pulmonary infection and respiratory failure; (4) Patients with malnutrition or excessive obesity; (5) Patients with the expected survival period <6 months; and (6) Patients with the training participation rate $<80\%$.

All 120 patients included in this study were aware of the purpose, significance, content, and confidentiality of this study and they signed their informed consent. They were equally divided into group p and group q according to the order of admission. There was no significant difference in general information between the two groups ($P > 0.05$), which can be used as the research subjects. See details in **Table 1**.

Standards of Withdrawal From Experiment

In the following conditions, the study group judged that the patients were unable to continue the experiment, and their case record forms were retained, but the data analysis was not performed. These conditions were: (1) Patients who had sudden exacerbation during the experiment and (2) In the experiment, subjects who were unwilling to continue the clinical trials, and who applied to the study group for withdrawal from the experiment.

Methods

All patients were given pulmonary rehabilitation and health care training. A team on rehabilitation and health care training was constructed, and the team members included the attending physicians, supervisor nurses, and primary nurses. After admission, the primary nurses collected the clinical information of patients, carried out the health promotion for patients, and improved their training compliance. The team on rehabilitation and health care training developed programs of the pulmonary rehabilitation's exercise training based on the actual circumstances of the patients. The patients were given deep pursed lips breathing at 1–2 weeks after discharge, with a daily training for 0.5 h. The deep pursed lips breathing meant that the patients inhaled slowly with the nose until the gas could not be inhaled, and then exhaled slowly with the mouth in contracted lips after 5 s. The relaxation and contraction exercise in muscles of the whole body was implemented in the third week after discharge, with daily training for 0.5 h. The 1–3 weeks of training

TABLE 1 | Comparison of patients' general information.

Groups	Group p ($n = 60$)	Group q ($n = 60$)	X^2/t	*P*
Male/female	32/28	34/26	0.135	0.714
Age (years)	52.75 ± 8.67	53.35 ± 8.40	0.385	0.701
Body mass (kg)	64.98 ± 2.41	64.77 ± 2.56	0.463	0.645
BMI (kg/m^2)	22.14 ± 1.65	22.10 ± 1.47	0.140	0.889
Pathological types			0.035	0.853
Squamous cell carcinoma	24	25		
Adenocarcinoma	36	35		
TNM staging			0.039	0.843
III	41	42		
IV	19	18		
Surgical procedures			0.035	0.853
Thoracic surgery	24	25		
Video-assisted thoracoscopic surgery	36	35		
Resection range				
Single lobectomy	25	24	0.035	0.853
Bilobectomy	15	16	0.044	0.835
Wedge resection	14	15	0.046	0.831
Sleeve resection	6	5	0.100	0.752
Educational level				
Primary school	12	14	0.196	0.658
Senior middle school	32	30	0.134	0.715
University	16	16	0.000	1.000
Monthly profit (yuan)			0.141	0.707
≥4000	22	24		
<4000	38	36		
Living habits				
Smoking history	40	38	0.147	0.702
Drinking history	32	30	0.134	0.715
Place of residence			0.034	0.854
Town	26	27		
Countryside	34	33		

programs were repeated in the fourth week, with daily training for 1 h. On Wednesday and Sunday, the centralized training was performed. At first, the patients were asked to repeat the deep pursed lips breathing, followed by progressive respiratory muscle contractions. After the patients took a semi-recumbent position, they took a deep breath slowly for three times and then held their breath, and they should contract the respiratory and abdominal muscles. When the patients could not persist, they could exhale quickly, and then gradually relax the muscle groups, and finally contract the muscles of the whole body. After the patients took a semi-recumbent position, the muscle contraction was performed from far to near and from top to bottom. Patients stretched themselves for 5 s with the taut muscles, and the muscles of the whole body were relaxed. In addition, the patients could jog according to the actual situation, first walking slowly at a speed of 30 steps/min, then walking fast at a speed of 75 steps/min, and finally jogging.

In addition, group p was treated with the osimertinib, and they were given 80 mg of osimertinib mesylate (AstraZeneca Pharmaceutical Co., Ltd.; specification: 80 mg; NMPA approval No.: J20180027) once a day. Group q received the treatment of pemetrexed and cisplatin. The 500 mg/m^2 of pemetrexed disodium (Sichuan Huiyu Pharmaceutical Co., Ltd.; specification: 500mg; NMPA approval No.: H20173301) was intravenously infused on the first day, and 75 mg/m^2 of cisplatin (Shandong Luoxin Pharmaceutical Group Co., Ltd.; specification: 10 mg; NMPA approval No.: H20046375) was intravenously infused on the second day with 3 weeks as 1 cycle. The routine examinations, such as blood routine, hepatorenal function, electrocardiogram, and dexamethasone (Guangdong Luofushan China National Medicines Co., Ltd.; specification: 0.75 mg; NMPA approval No.: H44024841) and folic acid tablets (Jiangxi Pharmaceutical Co., Ltd.; 0.4 mg; NMPA approval No.: H36020872) were given before chemotherapy.

Observational Criteria

General Information

The patients' data on social demography and clinic were collected including gender, age, body mass, BMI, pathological types, TNM staging (13), surgical procedures, resection range, educational level, monthly profit, living habits, and place of residence.

Pulmonary Function

The forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), FEV_1/FVC, the percentage of maximal ventilatory volume in predicted value (MVV%), and maximal vital capacity were recorded before treatment and at 1 month after treatment using a pulmonary function instrument [Body Plethysmograph, NMPA (I) 20172071024] in Medical Graphics Corporation.

Incidence of Complications

The types of complications were recorded, the number of patients with complications was counted, and the incidence of complications was calculated. The incidence of complications = (the number of patients with complications/total cases) ×100%.

QOL

The QOL instruments for patients with cancer: lung cancer (QLICP-LU) (14) developed by Wan Chonghua et al. was used to investigate patients' QOL before treatment and 1 month after treatment. The scale included the somatic function (7–35 points), psychological function (12–60 points), social function (6–30 points), common symptoms and adverse reactions (7–35 points), and specific modules (8–40 points) with the test-retest reliability as 0.78. The α value of internal reliability in various fields and the test-retest reliability were above 0.70, which had favorable reliability and validity.

Statistical Treatment

In this study, the experimental data were processed by SPSS20.0, and GraphPad Prism 7 (GraphPad Software, San Diego, USA) was used to draw pictures of the data. The items included in the

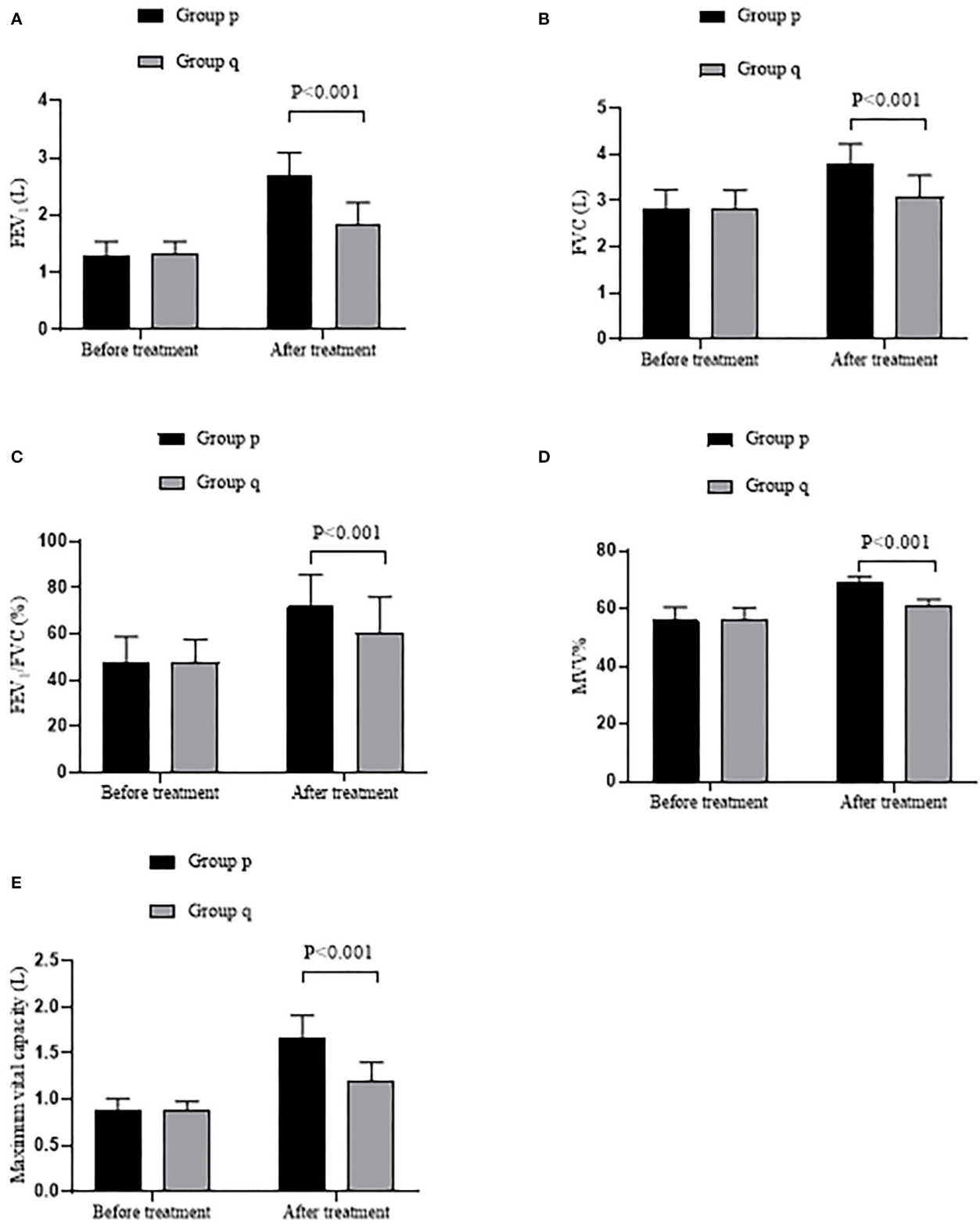


FIGURE 1 | Comparison of patients' pulmonary function (x ± s). **(A)** showed the FEV_1 (L). **(B)** showed the FVC (L). **(C)** showed the FEV_1/FVC (%). **(D)** showed the MVV%. **(E)** showed the maximal vital capacity (L).

TABLE 2 | Comparison of patients' incidence of complications [n (%)].

Groups	Group p ($n = 60$)	Group q ($n = 60$)	X^2	P
Pulmonary infection	0 (0.0)	1 (1.7)	1.008	0.315
Dyspnea	2 (3.3)	5 (8.3)	1.365	0.243
Atelectasis	2 (3.3)	4 (6.7)	0.702	0.402
Respiratory failure	0 (0.0)	1 (1.7)	1.008	0.315
Eating disorder	0 (0.0)	1 (1.7)	1.008	0.315
The number of occurrences	4 (6.7)	12 (20.0)	4.615	0.032

study were enumeration data and measurement data tested by x^2 and t–tests. When $p < 0.05$, the differences were considered to be statistically significant.

RESULTS

Comparison of Patients' General Information

There was no statistical difference in the general information between the two groups ($P > 0.05$). See details in **Table 1**.

Comparison of Patients' Pulmonary Function

Compared with group q, the pulmonary function was observably higher in group p after treatment ($P < 0.001$). See details in **Figure 1**.

Compared with group q, the FEV_1 level, FVC level, FEV_1/FVC level, MVV% level, and maximal vital capacity were observably higher in group p after treatment (2.68 ± 0.41 vs. 1.82 ± 0.40, 3.78 ± 0.45 vs. 3.08 ± 0.47, 71.82 ± 13.98 vs. 60.41 ± 15.75, 68.98 ± 2.10 vs. 60.77 ± 2.47, and 1.67 ± 0.24 vs. 1.21 ± 0.19, all $P < 0.001$), with no significant difference before treatment (1.30 ± 0.24 vs. 1.32 ± 0.22, 2.80 ± 0.44 vs. 2.81 ± 0.42, 47.73 ± 11.17 vs. 47.74 ± 9.96, 55.98 ± 4.50 vs. 55.90 ± 4.48, and 0.89 ± 0.12 vs.0.88 ± 0.10, all $P > 0.05$).

Comparison of Patients' Incidence of Complications

Compared with group q, the incidence of complications in group p was significantly lower ($P < 0.05$). See details in **Table 2**.

Comparison of Patients' QOL

Compared with group q, QOL scores in group p were observably higher after treatment ($P < 0.001$). See details in **Figure 2**.

Compared with group q, the somatic function score, psychological function score, social function score, the score of common symptoms and adverse reactions, and the score of specific modules were markedly higher in group p after treatment (26.20 ± 1.90 vs. 22.20 ± 1.91, 40.20 ± 1.84 vs. 31.27 ± 1.97, 24.17 ± 1.92 vs. 20.23 ± 1.89, 31.78 ± 1.66 vs. 26.22 ± 1.93, and 35.12 ± 1.70 vs. 30.18 ± 1.92, all $P < 0.001$), with no significant difference before treatment (20.17 ± 2 vs. 20.23 ± 1.95, 28.27 ± 1.85 vs. 28.25 ± 1.89, 20.18 ± 1.96 vs. 20.20 ± 1.95, 24.25 ± 1.93 vs. 24.18 ± 1.94, and 29.95 ± 1.93 vs. 29.58 ± 1.75, all $P > 0.05$).

DISCUSSION

According to the epidemiological data in 2018, the incidence and mortality of lung cancer ranked first in the malignant tumors in China. About 1.8 million patients died from lung cancer each year, accounting for 18.4% of all cancer death cases (15, 16). After the pathological types are clarified, the comprehensive treatment measures for patients can effectively reduce their mortality and prolong their survival period. The comprehensive treatment measures mainly include surgery, chemotherapy, and radiotherapy. Since SCLC often has metastasis in the early and middle stages, chemotherapy and radiotherapy were generally adopted in the clinic (17), while patients with NSCLC with a limited disease are often treated with surgery and postoperative radiotherapy and chemotherapy. Although radical resection of lung cancer can remove the lesions of lung cancer and reduce the clinical symptoms of patients with NSCLC, it cannot effectively improve the pulmonary function of patients (18, 19). In this study, patients were given pulmonary rehabilitation and health care training after the radical resection of lung cancer. The training methods mainly included subjective testing, breath instruction, and intensive training, which could regulate the respiratory contraction motion of patients to the maximum extent, improve the function of tracheal epithelium, and enhance the elimination effect of airway ciliary, thereby reducing the incidence of complications such as respiratory tract infection. In the process of the repeated exercises, patients could produce sufficient compensative capacity and enhance their lung capacity and lung function (19). Therefore, the pulmonary function indexes such as FEV_1 and FVC in the two groups after treatment were improved. Scholars Van Haran Robert M et al. have found that the pulmonary function indexes, such as FEV_1 and FVC of patients, were significantly improved by the application of the pulmonary rehabilitation and health care training in the rapid rehabilitation of lung cancer patients after surgery, indicating that the rehabilitation training method has a clear effect on lung cancer patients (20). However, the patients in group p had a more significant increase in pulmonary function compared with those in group q. The reason is that the pemetrexed used in group q has an anti-tumor effect, but the therapeutic effect on NSCLC is not obvious, while the osimertinib used in group p has a targeted therapeutic effect on the patients with NSCLC with EGFR gene mutation.

In general, postoperative chemotherapy for patients with SCLC can often obtain an ideal therapeutic effect, but the chemotherapy effect in patients with NSCLC is not very clear,; patients with the feeble constitution, poor bone marrow function, or dysfunction in vital organs are especially unsuitable for chemotherapy (21–24). Patients with NSCLC with the EGFR gene mutation even get worse after chemotherapy, so patients with NSCLC should be careful with chemotherapy and look for more effective treatment methods. Osimertinib, a third generation of EGFR-TKI, is the primary therapeutic drug after patients have drug resistance to the first and second generation of targeted drugs. Phase I and III clinical trials have shown that simertinib is effective in patients with locally advanced or metastatic EGFR mutation-positive non-small-cell lung cancer

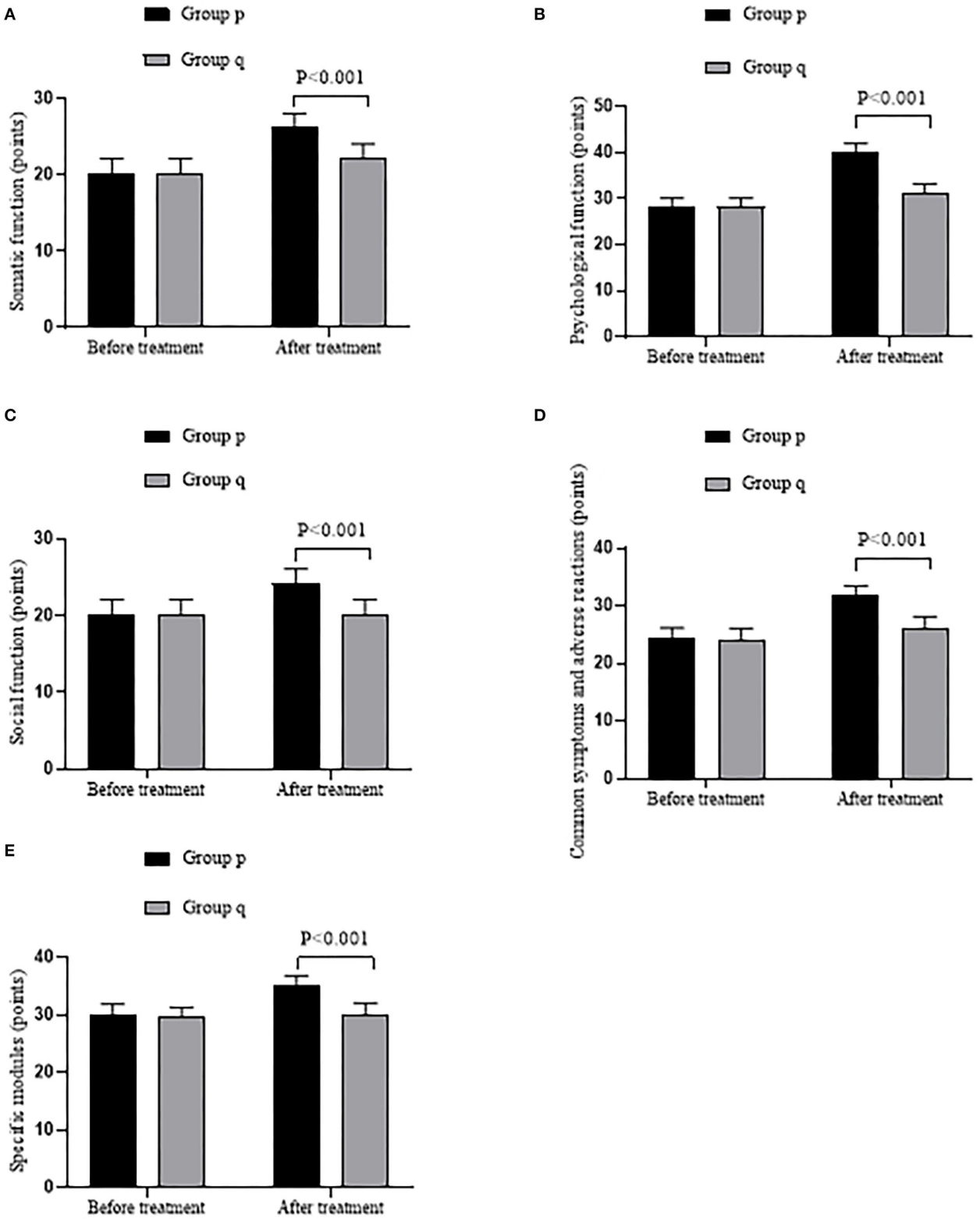


FIGURE 2 | Comparison of patient's QOL ($\bar{x} \pm s$, points). **(A)** showed the somatic function. **(B)** showed the psychological function. **(C)** showed the social function. **(D)** showed the common symptoms and adverse reactions. **(E)** showed the specific modules.

(25). After the previous EGFR-TKI treatment, patients in the disease progression received the simertinib, with a disease control rate of 92.0%. Therefore, osimertinib is certified as a ground-breaking therapy drug for NSCLC by the FDA (26, 27). The mechanism of osimertinib is that the acrylamide contained in osimertinib can form the covalent bonding with C797 at the edge of the ATP binding site in the catalytic domain of the EGFR gene, and then the irreversible binding occurs with the specific EGFR mutations, to solve the problem of abnormal signal transduction pathway in downstream cells caused by EGFR mutation, inhibit the excessive proliferation and mutation of tumor cells due to the EGFR mutation, and reduce the metastasis rate. The adverse reactions of osimertinib are mild compared with the conventional chemotherapeutic agents such as pemetrexed, and osimertinib shows a better tolerance in most clinical trials. Zhao Yang et al. have found that patients with osimertinib have no toxic reaction at a dose of 240 mg/d, and the incidence of severe adverse reactions is only 2%. Therefore, patients treated with osimertinib have a better organismal tolerance and an ideal long-term QOL (28). Chul et al. have confirmed that patients treated with osimertinib have significant progress in psychological and social functions, with a significant improvement in QOL (29), and the results are in accordance with this study. It is worth noting that the effect of osimertinib in improving the QOL of patients was mostly studied individually in current academia. This study paid attention to the combined application of osimertinib and other treatment measures, which found that osimertinib could improve the effect of pulmonary function training, and the higher QOL score in group p was consistent with the argument.

It is worth noting that long-term use of osimertinib will also produce drug resistance. No drug is available for patients with osimertinib resistance in the current practice. Scholars Xueting et al. have found that dihydroartemisinin can overcome the drug resistance (30), but there still lacks a large number of control studies. The fourth generation of EGFR-TKI drugs needs to be developed with osimertinib as a breakthrough point. At present, patients should be given the osimertinib combined with the pulmonary rehabilitation and health care training in the clinic, thereby providing a favorable physiological basis for the treatment of patients, reducing the postoperative morbidity, and improving their QOL.

In conclusion, the combination of osimertinib and pulmonary rehabilitation and health care training can improve the pulmonary function of patients with NSCLC with radical resection of lung cancer, and reduce the postoperative morbidity, thereby improving their QOL, which is conducive to reducing the patient's and society's medical burden.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First Affiliated Hospital of Zhengzhou University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HX and YY: conception, design, data analysis, and interpretation. All authors: administrative support, provision of study materials or patients, collection and assembly of data, manuscript writing, and final approval of manuscript.

REFERENCES

1. Che GW. Current situation and strategy of day surgery in patients with lung cancer by enhanced recovery after surgery. *Zhongguo Fei Ai Za Zhi*. (2022) 23:1–4. doi: 10.3779/j.issn.1009-3419.2020.01.01
2. Rogers LJ, Bleetman D, Messenger DE, Joshi NA, Wood L, Rasburn NJ, et al. The impact of enhanced recovery after surgery (ERAS) protocol compliance on morbidity from resection for primary lung cancer. *J Thorac Cardiovasc Surg*. (2018) 155:1843–52. doi: 10.1016/j.jtcvs.2017.10.151
3. Zheng Y, Mao M, Ji MF, Zheng QG, Liu L, Zhao ZG, et al. Does a pulmonary rehabilitation based ERAS program (PREP) affect pulmonary complication incidence, pulmonary function and quality of life after lung cancer surgery? Study protocol for a multicenter randomized controlled trial. *BMC Pulm Med*. (2020) 20:44. doi: 10.1186/s12890-020-1073-6
4. Hsieh MK, Bowles DR, Canseco JA, Sherman MB, Schroeder GD, Vaccaro AR. Is open surgery for metastatic spinal cord compression secondary to lung cancer really beneficial? A systematic review. *World Neurosurg*. (2020) 59:116–1210. doi: 10.1016/j.wneu.2020.08.098
5. Cecilia P, Finn MB, Ling CW, Nilanjan C, Emmanuel K, Richard M, et al. Poor preoperative quality of life predicts prolonged hospital stay after VATS lobectomy for lung cancer. *Eur J Cardiothorac Surg*. (2021) 59:116–21. doi: 10.1093/ejcts/ezaa245
6. Mendogni P, Mazzucco A, Palleschi A, Rosso L, Righi I, Carrinola R, et al. Uniportal and three-portal video-assisted thoracic surgery pulmonary lobectomy for early-stage lung cancer (UNIT trial): study protocol of a single-center randomized trial. *Trials*. (2021) 22:163. doi: 10.1186/s13063-021-05115-w
7. Li XH, Yu RF, Wang P, Wang AW, Huang HM. Effects of exercise training on cardiopulmonary function and quality of life in elderly patients with pulmonary fibrosis: a meta-analysis. *Int J Environ Res Public Health*. (2021) 18. doi: 10.3390/ijerph18147643
8. Mangiameli G, Cioffi U, Alloisio M, Testori A. Lung metastases: current surgical indications and new perspectives. *Front Surg*. (2022) 9:884915. doi: 10.3389/fsurg.2022.884915
9. Ding L, Li HY, Wang YP. Application of Jianpi Xiaoai recipe combined with cisplatin and adriamycin in the treatment of endometrial cancer and its effect on disease control rate. *Evid Based Complement Alternat Med*. (2021) 2021:2258183. doi: 10.1155/2021/2258183
10. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. (2013) 310:2191–4. doi: 10.1001/jama.2013.281053

11. Wang L. Chinese standard for diagnosis and treatment of primary lung cancer (2015 Edition). *Chin. J. Cancer*. (2016) 37:433–436.
12. Elkhayat H, Elsayed HH, Adel W, Elkhouly AG, Abdellateef A, Amin WAM, et al. Thoracic surgery in Egypt. *J Thorac Dis*. (2022) 14:1282–95. doi: 10.21037/jtd-21-723
13. Wang CM, Lai YT Li PF, Su JH, Che GW. Influence of enhanced recovery after surgery (ERAS) on patients receiving lung resection: a retrospective study of 1749 cases. *BMC Surg*. (2021) 21:115. doi: 10.1186/s12893-020-00960-z
14. Li ZF, Zhang BY, Yang FY, Yang LW. Pathological complete response to radical surgery after receiving durvalumab plus neoadjuvant chemotherapy for 1 limited-stage small cell lung cancer patient: a case report. *Transl Cancer Res*. (2022) 11:973–9. doi: 10.21037/tcr-22-729
15. Deng HS, Liu J, Cai XY, Chen JW, Gaetano R, René Horsleben P, et al. Radical minimally invasive surgery after immuno-chemotherapy in initially-unresectable stage IIIB non-small cell lung cancer. *Ann Surg*. (2022) 275:e600–2. doi: 10.1097/SLA.0000000000005233
16. Ke LH, Cui SP, Chen S, Hu B, Li H. Dynamics of D-dimer in non-small cell lung cancer patients receiving radical surgery and its association with postoperative venous thromboembolism. *Thorac Cancer*. (2020) 11:2483–92. doi: 10.1111/1759-7714.13559
17. Li RY, Wang K, Qu CH Qi WF, Fang T, Yue WM, Tian H. The effect of the enhanced recovery after surgery program on lung cancer surgery: a systematic review and meta-analysis. *J Thorac Dis*. (2021) 13:3566–86. doi: 10.21037/jtd-21-433
18. Higashimoto Y, Ando M, Sano A, Saeki S, Nishikawa Y, Fukuda K, et al. Effect of pulmonary rehabilitation programs including lower limb endurance training on dyspnea in stable COPD: a systematic review and meta-analysis. *Respir Investig*. (2020) 58:355–66. doi: 10.1016/j.resinv.2020.05.010
19. Shiyko MP, Brian S, Greene PB, Joshua S, Burkhalter JE. Intra-individual study of mindfulness: ecological momentary perspective in post-surgical lung cancer patients. *J Behav Med*. (2019) 42:102–10. doi: 10.1007/s10865-018-9942-7
20. Haren V, Robert M, Reza JM, Gabriel EM, Arlene MC, Mara BA, et al. Enhanced recovery decreases pulmonary and cardiac complications after thoracotomy for lung cancer. *Ann Thorac Surg*. (2018) 106:272–9. doi: 10.1016/j.athoracsur.2018.01.088
21. Esra CH, Arzu T. How does health education given to lung cancer patients before thoracotomy affect pain, anxiety, and respiratory functions? *J Cancer Educ*. (2019) 34:966–72. doi: 10.1007/s13187-018-1401-1
22. Marfil-Sánchez A, Seelbinder B, Ni Y, Varga J, Berta J, Hollosi V, et al. Gut microbiome functionality might be associated with exercise tolerance and recurrence of resected early-stage lung cancer patients. *PLoS ONE*. (2021) 16:e0259898. doi: 10.1371/journal.pone.0259898
23. Ding LL, Xu ZY, Zhao ZY Li HP, Xu AH. Effects of pulmonary rehabilitation training based on WeChat App on pulmonary function, adverse mood and quality of life of COVID-19 patients: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. (2021) 100:e26813. doi: 10.1097/MD.0000000000026813
24. Kong M, Zheng HK, Ding LL, Jin K, Shen JF, Ye MH, et al. Perioperative pulmonary rehabilitation training (PPRT) can reduce the cost of medical resources in patients undergoing thoracoscopic lung cancer resection: a retrospective study. *Ann Palliat Med*. (2021) 10:4418–27. doi: 10.21037/apm-21-478
25. Yasuda H, Ichihara E, Sakakibara-Konishi J, Zenke Y, Takeuchi S, Morise M, et al. A phase I/II study of osimertinib in EGFR exon 20 insertion mutation-positive non-small cell lung cancer. *Lung Cancer*. (2021) 162:140–6. doi: 10.1016/j.lungcan.2021.10.006
26. Lai L, Shen QP, Wang YJ, Chen LT, Lai JJ, Wu ZB, et al. Polyphyllin I reverses the resistance of osimertinib in non-small cell lung cancer cell through regulation of PI3K/Akt signaling. *Toxicol Appl Pharmacol*. (2021) 419:115518. doi: 10.1016/j.taap.2021.115518
27. Provencio M, Terrasa J, Garrido P, Campelo RG, Aparisi F, Diz P, et al. Osimertinib in advanced EGFR-T790M mutation-positive non-small cell lung cancer patients treated within the Special Use Medication Program in Spain: OSIREX-Spanish Lung Cancer Group. *BMC Cancer*. (2021) 21:230. doi: 10.1186/s12885-021-07922-5
28. Yang Z, Shuyan L, Xi Y, Li C, Shengping W, Tong T, et al. Overall survival benefit of osimertinib and clinical value of upfront cranial local therapy in untreated EGFR-mutant nonsmall cell lung cancer with brain metastasis. *Int J Cancer*. (2022) 150: 1318-1328. doi: 10.1002/ijc.33904
29. Cho BC, Felip E, Hayashi H, Thomas M, Lu S, Besse B, et al. MARIPOSA: phase 3 study of first-line amivantamab + lazertinib versus osimertinib in EGFR-mutant non-small-cell lung cancer. *Future Oncol*. (2022) 18:639–47. doi: 10.2217/fon-2021-0923
30. Cai XT, Miao J, Sun RW, Wang SN, Molina-Vila MA, Chaib I, et al. Dihydroartemisinin overcomes the resistance to osimertinib in EGFR-mutant non-small-cell lung cancer. *Pharmacol Res*. (2021) 170:105701. doi: 10.1016/j.phrs.2021.105701

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xu, Guo and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 13 June 2022
doi: 10.3389/fonc.2022.834028

Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images

Ahila A[1], Poongodi M[2]*, Sami Bourouis[3], Shahab S. Band[4]*, Amir Mosavi[5]*, Shweta Agrawal[6] and Mounir Hamdi[2]*

[1] Department of Electronics and Communication Engineering, Sethu Institute of Technology, Kariapatti, India, [2] College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar, [3] Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, [4] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliou, Taiwan, [5] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [6] IAC, SAGE University, Indore, India

OPEN ACCESS

Edited by:
Celestine Iwendi,
University of Bolton, United Kingdom

Reviewed by:
Senthilkumar Mohan,
VIT University, India
Noble Anumbe,
University of South Carolina,
United States

***Correspondence:**
Ahila A
akhilaamarnath27@gmail.com
Poongodi M.
dr.m.poongodi@gmail.com
Shahab S. Band
shamshirbands@yuntech.edu.tw
Amir Mosavi
amir.mosavi@kvk.uni-obuda.hu

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 12 December 2021
Accepted: 14 March 2022
Published: 13 June 2022

Citation:
A A, M P, Bourouis S, Band SS, Mosavi A, Agrawal S and Hamdi M (2022) Meta-Heuristic Algorithm-Tuned Neural Network for Breast Cancer Diagnosis Using Ultrasound Images. Front. Oncol. 12:834028. doi: 10.3389/fonc.2022.834028

Breast cancer is the most menacing cancer among all types of cancer in women around the globe. Early diagnosis is the only way to increase the treatment options which then decreases the death rate and increases the chance of survival in patients. However, it is a challenging task to differentiate abnormal breast tissues from normal tissues because of their structure and unclear boundaries. Therefore, early and accurate diagnosis and classification of breast lesions into malignant or benign lesions is an active domain of research. Over the decade, numerous artificial neural network (ANN)-based techniques were adopted in order to diagnose and classify breast cancer due to the unique characteristics of learning key features from complex data *via* a training process. However, these schemes have limitations like slow convergence and longer training time. To address the above mentioned issues, this paper employs a meta-heuristic algorithm for tuning the parameters of the neural network. The main novelty of this work is the computer-aided diagnosis scheme for detecting abnormalities in breast ultrasound images by integrating a wavelet neural network (WNN) and the grey wolf optimization (GWO) algorithm. Here, breast ultrasound (US) images are preprocessed with a sigmoid filter followed by interference-based despeckling and then by anisotropic diffusion. The automatic segmentation algorithm is adopted to extract the region of interest, and subsequently morphological and texture features are computed. Finally, the GWO-tuned WNN is exploited to accomplish the classification task. The classification performance of the proposed scheme is validated on 346 ultrasound images. Efficiency of the proposed methodology is evaluated by computing the confusion matrix and receiver operating characteristic (ROC) curve. Numerical analysis revealed that the proposed work can yield higher classification accuracy when compared to the

prevailing methods and thereby proves its potential in effective breast tumor detection and classification. The proposed GWO-WNN method (98%) gives better accuracy than other methods like SOM-SVM (87.5), LOFA-SVM (93.62%), MBA-RF (96.85%), and BAS-BPNN (96.3%)

Keywords: breast cancer detection, computer-aided diagnosis, supervised learning, texture features, ultrasound imaging, wavelet neural network

1 INTRODUCTION

Breast cancer is one among the most prevailing cancers diagnosed in women across the globe. It can occur in both men and women, but is more prevalent in women. According to the American Cancer Society (ACS) and World Health Organization (WHO), around 1.7 million people are diagnosed with breast cancer every year (1). Breast cancer can be treated and restored if it is detected at an early stage. Therefore, early detection of breast cancer is of prime importance for increasing the survival rate of patients (2). Screening is the best way of reducing mortality rate from breast cancer. Medical images are crucial sources of useful information. Several imaging techniques including mammography, ultrasound (US) imaging, and elastography are being widely used for diagnosing breast cancer (3, 4). Mammography is the most popular imaging technique for detecting cancer. However, it has some limitations like low sensitivity over ionizing radiations and is not suitable for dense breasts (5–8). Therefore, US or sonography is a valuable tool over mammography owing to its speed, inexpensiveness, and robustness (9), thereby it is safe and well tolerated by patients. Furthermore, US does not use any harmful radiation and has the ability to detect and differentiate benign from malignant tumors. Sonogram images are visualized, analyzed, and diagnosed manually by the physician or radiologist, but these are relatively tedious and time-consuming processes (10). In addition to this, manual analysis relies upon physician experience. Even experienced physicians may follow different diagnosis methods for manual analysis and have a sensitivity of 84% and specificity of 91% in breast cancer diagnosis. Therefore, there is a requirement to propose an efficient computer-aided diagnosis (CAD) system for US image analysis to enhance the diagnosis accuracy and help physicians make accurate timely decisions. An automated CAD system can help the physician to analyze the US image precisely and emphasize the suspected areas that need further investigation. Numerous methods for breast cancer diagnosis and developing sophisticated methods for enhancing the diagnosis accuracy have been carried out in the past years (11). A CAD involves five major phases: preprocessing, segmentation, feature extraction, feature selection, and classification (12). formulated a hybrid method for breast cancer diagnosis by integrating particle swarm optimization (PSO) and support vector machine (SVM) algorithms (13, 14).

PSO is mainly adopted for feature selection. The multiple instances learning-based breast cancer classification method is investigated by (15). This method is based on extracting the features from enhanced images and adopting SVM for segmentation (16). A self-organizing map (SOM) is used for converting the instance space to concept space. Finally, SVM is used for diagnosing breast cancer (17). proposed a classification approach for breast cancer diagnosis. Here, the US images are preprocessed with histogram equalization and the gamma correction method. Subsequently, texture features are extracted and classified as benign or malignant with the SVM classifier (18). reduced the speckle noise from breast US images by using an enhanced oriented despeckling anisotropic diffusion bilateral filter (19, 20). developed a method for increasing the classification accuracy. Here, the input images are preprocessed and segmented using the watershed algorithm. Morphological and texture features are obtained and fed as input to the classifier (21). employed an ANN for formulating the breast cancer diagnosis system. They computed feature scores based on breast imaging reporting and data system (BIRDS) criteria and then they used the biclustering method by (22, 23) to generate the hidden features. A hybrid model based on the fruit fly optimization algorithm (FLA) and SVM was assessed by (24). They employed FLA for tuning SVM parameters (25). The proposed particle swarm optimization, non-domination sorting with classifier technique model, and Bayes' theorem help to select the most significant parameters and achieve optimum accuracy in breast cancer prediction (26). Jain et al. (2021) (27) proposes ensemble learning methods while making predictions. Bayes, K-nearest neighbors, AdaBoost, Gradient boosting, XGBoost, Random forest, ensembles, and neural networks are used for analyzing and predicting the dataset. It gives 87.248% and 87.7934% validation accuracy for predicting SARS79 CoV and SARS-CoV-2 virus respectively. The proposed algorithm produced the most accurate results, with 0.919 AUC and 87.248 percent validation accuracy for predicting SARS-CoV and 0.923 AUC and 87.7934 percent validation accuracy for predicting SARS-CoV-2 virus. Gupta et al. (2021) (28) proposed multimodal textual data consists of different emotions have been handpicked through the real time tweets related to COVID-19.After evaluating the most frequent word,modal vector has been created by using term frequency-inverse document frequency. Results showed that FLA-SVM achieved better classification accuracy. However, most of researchers have used machine learning models to perform classification tasks. Though machine learning is a powerful tool for image classification, it has some limitations like overfitting, increased learning time, and the fact that developing an optimal network structure is a tedious task and requires more time. Furthermore, optimizing parameters of the network is a difficult task. Parameter optimization is usually accomplished by experimentation. The major intention of this research is to

propose a CAD system to diagnose breast cancer in an automated manner so as to overcome the limitations of the aforementioned existing techniques given in the literature. The major contributions of this work are as follows:

(1) To develop an automated CAD system by adopting a meta-heuristic algorithm and machine learning (2), to optimize the wavelet neural network with the grey wolf optimization algorithm to improve the classification accuracy in US breast cancer diagnosis, and (3) to compare the obtained outcomes with prevailing methods so as to project the preeminence of the proposed CAD system. The paper is structured as follows. Section 2 demonstrates the proposed CAD system for breast cancer. Section 3 provides experimental results and comparison with the existing methods. Section 4 concludes the paper.

2 PROPOSED CAD SYSTEM

Research studies proved that ANN is an effective tool for breast cancer identification. Many researchers have employed ANN for developing a CAD system and showed comparatively better results. However, determining the optimal topology of ANN is a very difficult task. The number of hidden layers, hidden neurons, weights, and bias is usually determined by the trial-and-error method. This is a time-consuming process. To deal with this issue, bio-inspired algorithms like the genetic algorithm (GA), particle swarm optimization, and the fractional lion algorithm have been used for optimizing ANN parameters. The basic articulation of the suggested CAD system has been depicted in **Figure 1**.

2.1 US Image Preprocessing

Preprocessing is a crucial process in medical image processing. The aim of preprocessing breast US images is to improve the image quality and to remove unwanted information. **Table 1** shows the symbol's nomenclature. In this research work, preprocessing involves two processes: enhancement and despeckling. A sigmoid filter is applied to breast US images to improve the quality. Let an input breast US image be assumed as *H*. Mathematically, the sigmoid filter can be expressed as Equation 1:

$$H_{enh} = (H_{mx} - H_{mn}) \times \frac{1}{1 + e^{\left[\frac{A-a}{\gamma}\right]}} + H_{mn} \quad (1)$$

Where, *H* is an input image, H_{enh} is an enhanced image, mn is the minimum gray value, mx is the maximum value, and α and γ represent the center and width, respectively. In this paper, *lpha* = 4 and γ = 7 is used. **Figure 2** shows the processed images after enhancement.

After enhancement, interference-based despeckling followed by anisotropic diffusion (IDAD) is applied to the enhanced image to remove speckle noise. The bright and dark regions of the US image represent constructive and destructive interference, respectively. The IDAD filter will suppress only destructive interference without losing any image information. The

TABLE 1 | Nomenclature.

S.No	Symbol	Description
1	X	Input network
2	Y	Output network
3	H	Input breast US image
4	mn	Minimum gray value of the image
5	mx	Maximum gray value of the image
6	W	Weights between the layers
7	Ψ	Wavelet activation function
8	α	Center of the image
9	γ	Width of the image
10	*i*	Wavelet number
11	j	Wavelet level
12	k	Translation parameter
13	F	No. of features in the network
14	y	No. of classes in the network
15	Z	No. of hidden layers in the network
16	b	Bias
17	D	Modified distance vectors
18	C,E	GWO parameters

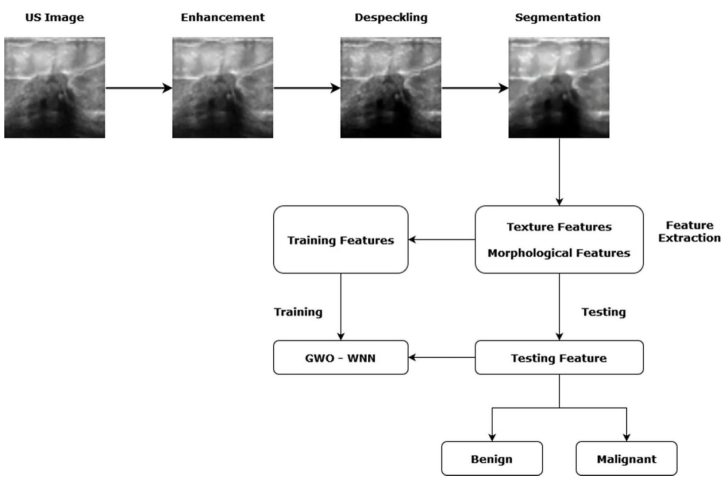


FIGURE 1 | Articulation of the Proposed CAD system.

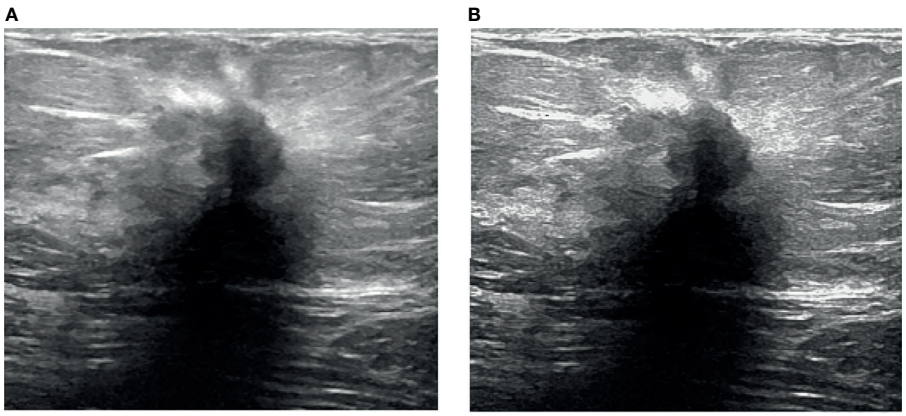


FIGURE 2 | Processing results. **(A)** Ultrasound image and **(B)** enhanced ultrasound image.

function of IDAD can be divided into two processes: median filter and interference suppression. A median filter with a kernel size of $n \in n$ is applied to the enhanced image. Image pixels are filtered with an $n \in n$ kernel around each pixel, replacing the central pixel value with the central value of all the pixels present in the kernel, and are mathematically represented in Equation 2.

$$H_{\text{med}}(X,Y) = \underset{i,j \in k_{n,n}}{\text{med}} \left[k_{n,n}(i,j) \otimes H_{\text{enh}}[X+i, Y+j] \right] \quad (2)$$

To prevent over-filtering, the kernel size was set to 3. Destructive interference suppression is expressed in Equation 3.

$$H_c(X,Y) = \max\left(H_{\text{enh}}(X,Y), H_{\text{med}}(X,Y)\right) \quad (3)$$

Where, A_c denotes the constructive interference image. Constructive interference suppression done by applying the median filter is applied to H_c to remove the remaining extreme single pixels. **Figure 3** shows the breast US image after despeckling. It was clearly observed that the IDAD filter preserves the edges and boundaries of the image.

2.2 Segmentation

Image segmentation plays a significant role in breast cancer diagnosis. It is mainly employed to isolate the lesion region from its background. In recent years, clustering methods such as fuzzy C-means (FCM) and k-means algorithms were adopted for image segmentation. The FCM algorithm can separate lesions precisely. This algorithm is sensitive to noise and other artifacts. On the other hand, the k-means algorithm is faster than FCM but it has some limitations like incomplete detection of malignant lesions. This research presents an automatic segmentation method by combining the FCM and k-means algorithms in order to obtain accurate segmentation. The despeckled image is used as input in the segmentation process. The proposed segmentation method starts with initialization of the number of clusters, M, maximum iterations, and the termination parameter. The cluster centers can be mathematically expressed in the following Equations 4 and 5:

$$\text{Initial mean} = \frac{(1:M) * n}{M+1} \quad (4)$$

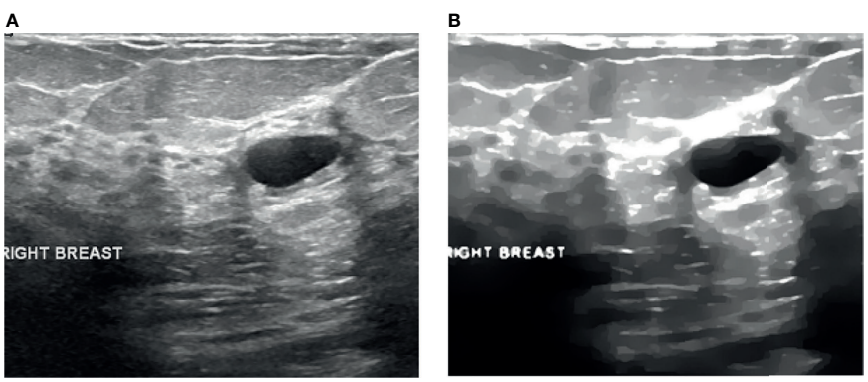


FIGURE 3 | Despeckling results. **(A)** Raw US image and **(B)** filtered image.

and,

$$n = \max(US\ \text{image}) + 1 \quad (5)$$

We assign each pixel to the nearest cluster center according to the minimum distance by examining the distance between the cluster centers and pixel value, and then recalculate the new cluster centers. This process repeats until the termination criterion is met. Both soft and hard clustering are used in the cluster formation. Hard clustering assigns each point based on the nearest cluster, in contrast, soft clustering assigns every point based on the degree of membership value. The clustered image is converted into a binary image with the thresholding method. Image binarization is one of the essential processes in image segmentation. The output of the binarization is the segmented image with a dark background and white foreground (lesion region). Active contour methods have been employed for segmentation and boundary tracking. In this research, active contour by the level set method is applied to track the boundary of the segmented lesions. The active contour by the level set method starts with initial boundary shapes represented as contours and is iteratively changed by executing shrinkage or expansion based on the constraints. The main advantage of active contour by level set is dividing the image into sub-images with continuous boundaries. A sample outcome of the segmentation process is shown in **Figure 4**.

2.3 Feature Extraction and Selection

The crucial steps in breast cancer classification are feature extraction and feature selection. It is well known that benign and malignant lesions differ in both shape and pixel values. It is important to select the features related to shape and boundary of the lesion. Morphological features depict the characteristics such as shape and border of the lesion. Benign tumor margins are smoother, round or oval-shaped, and regular, whereas, malignant margins are angular, irregular shaped, and edged. It is also important to consider texture features for breast cancer classification since there are differences in boundary and internal echoes of malignant and benign lesions in US images. Therefore, the proposed CAD system employs two features to conduct differentiation amid benign and malignant lesions. Various texture and morphological features that were selected for conducing this experiment are provided in **Figure 5**. The obtained morphological and texture features are combined in order to obtain a single feature set. Principal component analysis (PCA) has been adopted to reduce the size of the combined features so as to speed up the training and testing processes, and also to improve the classification rate.

2.4 Classification

Classification has been identified to be applicable in different domains including medical image classification. The classification task largely relies on how efficiently the discriminant features and classifiers were selected. In recent years, numerous computational intelligence (CI) techniques such as the feed forward neural network, SVM, and wavelet neural network (WNN) have been used for image classification due to their ability to learn and find the relation between input and output variables *via* a training process. However, the feed forward neural network is largely popular and the most commonly used to diagnose medical images. Wavelet is a mathematical tool used to analyze a signal or an image in different resolutions. It offers the potential of obtaining spatial and frequency domain features of a signal or an image. This paper develops a classifier by combining a neural network and the wavelet theory WNN to obtain a higher classification rate when compared to its peers. The peculiar advantages in using WNN when compared to multilayer perceptron (MLP) and SVM include the fact that WNN can handle non-linear, noisy, and periodic signals. Thus, it works well in determining the abnormalities in breast US images. Moreover, WNNs are robust and require only smaller training samples for tuning the parameters when compared with MLP and SVM. Also, WNNs offer faster convergence when compared with MLP and SVM. Hence, this proposed method employs WNN owing to these aforementioned enhanced features.

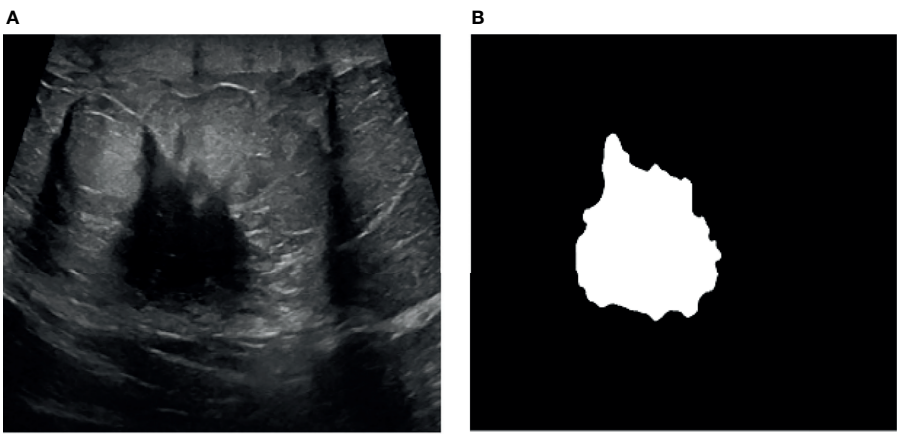


FIGURE 4 | ROI segmentation. **(A)** Ultrasound image and **(B)** segmented image.

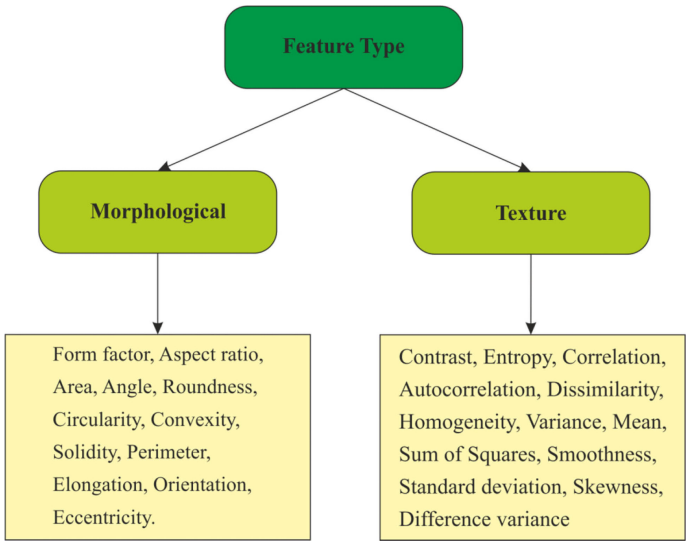


FIGURE 5 | Selected features for this proposed scheme.

Figure 6 shows the general structure of WNN. A typical WNN has three layers which are the input, hidden, and output layers. The input layer is responsible for receiving input signals or feature vectors from the source. The hidden layer has hidden neurons similar to neurons in standard neural networks (NNs). In standard NNs, the sigmoidal activation function is utilized at the hidden layer. Wavelets are subsequently incorporated as the activation function in the hidden layer of WNNs. Finally, the output layer corresponds to the linear combination of the wavelet basis function. The error between predicted values and expected values is calculated by cost function, and also represents itself in the format of a single real number. Based on the nature of problems, the cost function could be framed in various ways. There is no standard rule for selecting the mother wavelet function as the activation function.

In general, there are numerous categories or families of wavelets such as Daubechies, Symlets, Coiflets, biorthogonal, Haar wavelets, etc. In this paper, the Haar wavelet is employed owing to its characteristics toward simple structure and increased memory efficiency. The Levenbergmarquart (LM) algorithm is employed for minimizing the error and also to update the weights of WNN. It is noted that psi signifies the wavelet activation function. In a standard neural network, the tan sigmoidal activation function is used at the hidden layer. But, in the case of a wavelet neural network, wavelets are employed as

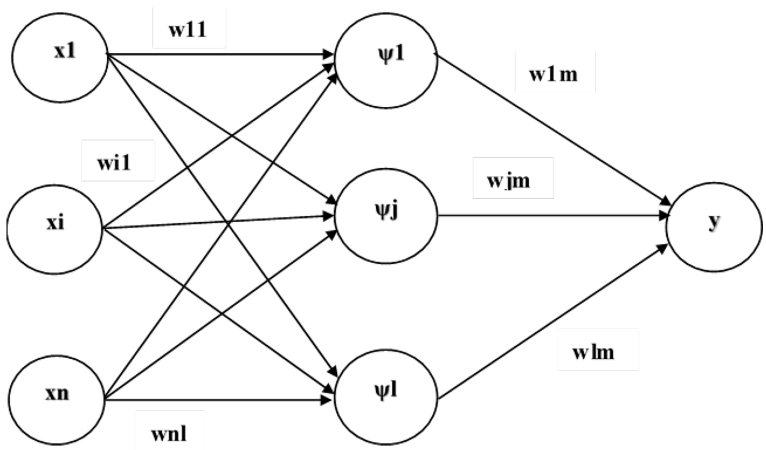


FIGURE 6 | Structure of WNN.

the activation function, thus, psi corresponds to the mother wavelet.

Let the input signal be represented as X={x_1, x_2, x_3,…,x_n} and the output signal be signified as Y={1, 2, 3, …, m}, with w_{ij} and w_{jk} corresponding to the weights between the layers. Ψ is the wavelet activation function. The weighted summation of the wavelet basis function at the hidden layers is closely related to the output layer. The output of the network with respect to the corresponding Haar wavelet function can be represented in the following Equations 6 and 7:

$$Yp = \sum_{j=1}^{I} W_{jk} \Psi ab \left[\sum_{i=1}^{n} X_{ij} W_{ij} + bi_j \right] \tag{6}$$

$$\Psi_{ab}\left(net_j\right) = \Psi \left[\frac{net_j - b_j}{a_j} \right] \tag{7}$$

Where a and b are the wavelet scale and translation parameter of the j_{th} hidden neuron. Although WNN is a promising tool for pattern recognition, optimizing the parameters of WNN like learning rate, weight, number of hidden neurons, etc., are quite difficult. The features of the Haar wavelet function that was employed in this proposed method are described below.

Let us assume, m = 2, k = 0,1,1,3,… $.m - 1, j$ = 0,1,2,3, and $i = m + k + 1$. Where i signifies the wavelet numbers, j corresponds to the wavelet levels, J signifies the highest resolution, and k corresponds to the translation parameter. The highest value of i is selected as i=2M and the lowest value is selected as i=2. The typical Haar wavelet families corresponding to $t \in [A,B]$ shall be mathematically expressed in the following Equations 8 and 9,

$$h_i(t) = \begin{matrix} 1 \text{ for } t \in |\xi_1(i), \xi_2(i)| \\ \& \\ -1 \text{ for } t \in |\xi_2(i), \xi_3(i)| \\ 0 \text{ for otherwise} \end{matrix} \tag{8}$$

Here,

$\xi_1(i) = A + 2k\mu\Delta x$, $\xi_2(i) = A + (2k+1)\mu\Delta x$, and $\xi_3(i) = A + (2k+1)\mu\Delta x$

$$\text{Also, } \mu = \frac{M}{m} \quad \Delta x = \frac{B-A}{2M} \tag{9}$$

The scaling function $h_{1(t)}$ shall be mathematically signified in Equation 10:

$$h_1(t) = \begin{cases} 1 \text{ for } t \in [A,B] \\ 0 \text{ for otherwise} \end{cases} \tag{10}$$

In general, the Haar wavelets are orthogonal with respect to each other. Hence, the orthogonality property shall be specified in Equation 11:

$$\int_A^B h_i(t) \times h_s(t) = 2^i \delta_{ij} = \begin{cases} 2^j \text{ i = s = } 2^j + k \\ 0 \qquad \text{i} \neq \text{S} \end{cases} \tag{11}$$

The orthogonality property of this Haar wavelet is mainly employed to enhance the efficacy of the present method. The basic function y(t) within the interval [A,B] shall be extended to the Haar wavelet and this wavelet extension shall be expressed in Equation 12:

$$y(t) = \sum_{i=1}^{2M} a_i h_i(t) \quad a_j = 2^j \int_A^B y(t) \times h_i(rmt)dt \tag{12}$$

To address the aforementioned limitations, the GWO algorithm is employed for tuning the WNN parameters. GWO is a category of nature-inspired algorithms motivated by the hierarchical organization and hunting attitude of grey wolves. GWO was initially formulated by (29). All grey wolves like to live as a group, thereby this GWO optimization algorithm possesses a social dominant hierarchy.

In a GWO algorithm, α represents the leader wolf since his/her order should be obeyed by the group. Here, α is responsible for making the decision regarding hunting, sleeping place, and so on. The β wolf should respect the leader, but commands δ and ω. Delta wolves should respect α and β, but dominates ω. The ω wolves have to submit to α, β, and δ. It was observed that the GWO algorithm possesses good exploration and exploitation characteristics when compared to particle swarm optimization, the genetic algorithm, and the cat swarm optimization algorithms. The corresponding pseudo code for the GWO algorithm which was employed in this experiment is given in the following Equations 13-23.

Initialize search agents, S randomly, variables v, and maximum iterations, I_{max}, a, C, and E.

$$D^U = \left| E^u X^u p(t) - X^U(t) \right| \tag{13}$$

$$X^u(t+1) = X^u p(t) - C^u \cdot D^u \tag{14}$$

$$C^u = 2.a \,.\, r1 - a \tag{15}$$

$$E^u = 2.r2 \tag{16}$$

Evaluate the fitness values of each agent.

X_α – First prime agent,
X_β – Second prime agent ,
X_δ -Third prime agent

$$D_a^U = E_1^u \,.\, X_\alpha^u - X^u \tag{17}$$

$$D_\beta^U = E_2^u \,.\, X_\beta^u - X^u \tag{18}$$

$$D_\delta^U = E_1^u \,.\, X_\delta^u - X^u \tag{19}$$

$$X_1^u = X_\alpha^u - C_1^u \cdot (D_\alpha^u) \tag{20}$$

$$X_2^u = X_\beta^u - C_1^u \cdot \left(D_\beta^u\right) \tag{21}$$

$$X_3^u = X_\delta^u - C_1^u \cdot \left(D_\gamma^u\right) \tag{22}$$

While (il_{max})for j = 1:S
Update position

$$X^u(t+1) = \frac{X_1^u + X_2^u + X_3^u}{3} \quad (23)$$

End for
Update a,C,E
Compute the quality of each search agent.
Update X_α, X_β, X_δ
i=i+1
End while
return X_α

In this proposed method, GWO-WNN is employed for developing WNN and is tuned with the best topology by tuning the weights, number of hidden layers and neurons, and respective learning rate. The algorithmic steps used to train and classify the breast US images are given in the following steps

Step 1: Read breast US image, *H*.

Step 2: Convert 3D image to 2D image.

Step 3: Enhance the quality of the image using Equation (24).

Step 4: Despeckle the enhanced image using Equation (26).

Step 5: Perform segmentation.

Step 6: Extract the features listed in **Figure 5**.

Step 7: Design a classifier.

Fi represents the count of features, y_o corresponds to the count of classes, and Z_n represents the count of hidden layers. The GWO algorithm determines the count of hidden neurons in each hidden layer weight and learning rate. Let g be the transfer function and it can be expressed in the following Equations 24 and 25.

$$Yp = g(Zj) \quad (24)$$

$$Z_j = \Psi_{ab} \sum_{j-1}^{n} X_i W_{ij} + b_j \quad (25)$$

$Z_p^j = \sum_v W_{v\cdot p} y_v^j + B_p$ Where, j represents the j^{th} hidden neuron, y_p is the output of the p^{th} neuron, $W_{i,j}$ is the weight, and b denotes the bias.

In this research work, the hidden layer and output layer used the wavelet activation function and linear activation function accordingly. Mean square error (MSE) is utilized as the fitness function for training, MSE can be expressed in Equation 26:

$$\mathrm{MSE} = \sum_{j \in t\mathrm{p}=1}^{\mathrm{Y}_\theta} \sum \left(\mathrm{t}_\mathrm{p}^\mathrm{j} - \mathrm{y}_\mathrm{p}^\mathrm{j}\right)^2 \quad (26)$$

Step 8: Optimize the parameter of WNN

Where t_p^j is the labeled class and y_p^j is the output obtained from the network.

3 RESULTS AND DISCUSSION

3.1 US Image Dataset

Breast US images were acquired from 346 patients from the database given below (https://scholar.cu.edu.eg/q=afahmy/pages/dataset) (30) including 249 malignant and 97 benign samples employed for assessing the effectiveness of the proposed GWO-tuned WNN. The entire dataset was divided into two sets: 70% of the data were utilized for training and the rest of the 30% were used for testing the classification ability of the system. A 10-cross fold validation was considered to test the efficiency of the presented system. The number of benign and malignant cases used in this research work is graphically displayed in **Figure 7**. The proposed scheme was implemented using MATLAB so as to analyze the efficacy of the presented system when compared to its peers.

3.2 Evaluation Indices

This section describes the evaluation parameters that were employed to analyze the efficacy of the presented scheme. Accuracy of the classification, specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), confusion matrix, and ROC were employed to analyze the classification performance.

3.3 Performance Evaluation

Figure 8 illustrates a variety of benign and malignant breast ultrasound pictures. The preprocessed 257 benign and malignant US pictures are shown in **Figures 9** and **10**, respectively. **Figure 11** depicts the confusion matrix to discriminate the cancer (malignant) from non-cancer (benign) breast lesions. In this research work, malignant tumors were defined as true positive and benign as true negative. From **Figure 11**, it can be observed that 245 out of 249 images were correctly diagnosed as malignant tumors. Likewise, 94 out of 97 images were correctly identified as benign. Consequently, true positive, false positive, false negative, and true negative values were obtained at 70.8%, 1.2, 0.9%, and 27.2%, respectively. The results exhibit superior classification accuracy of 98% with a sensitivity of 98.8%, specificity of 95.9%, PPV of 98.4%, and NPV of 96.9%.

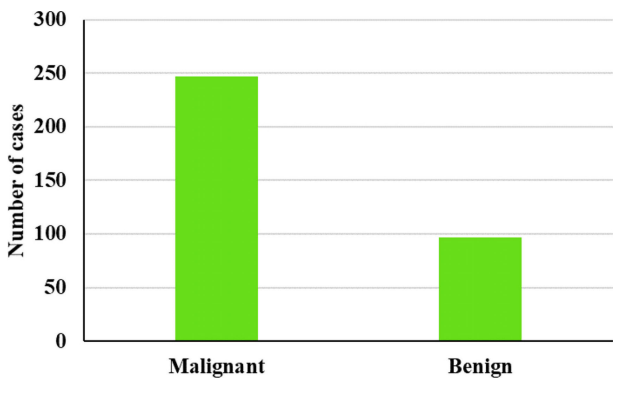


FIGURE 7 | US image dataset.

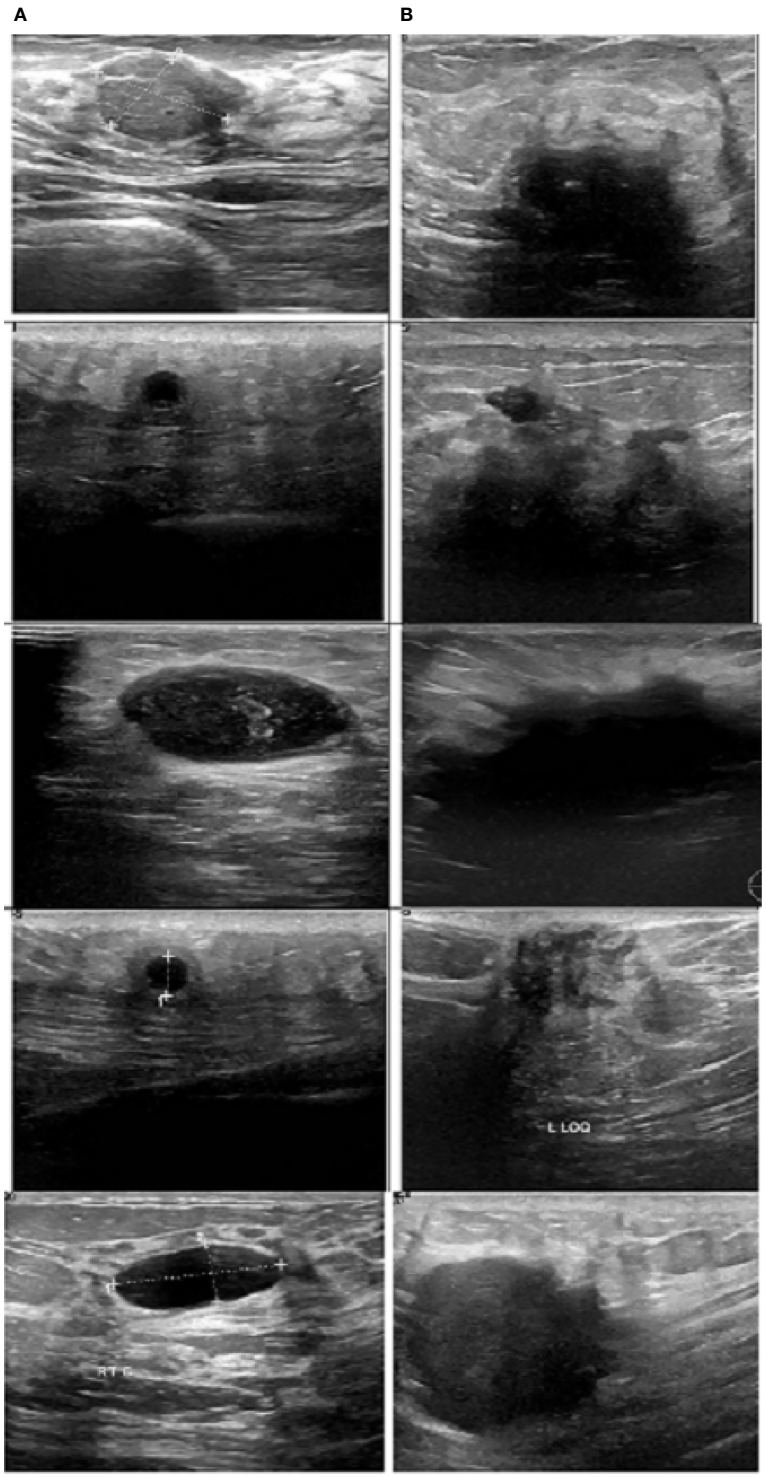


FIGURE 8 | Sample images. **(A)** Benign and **(B)** malignant (30)

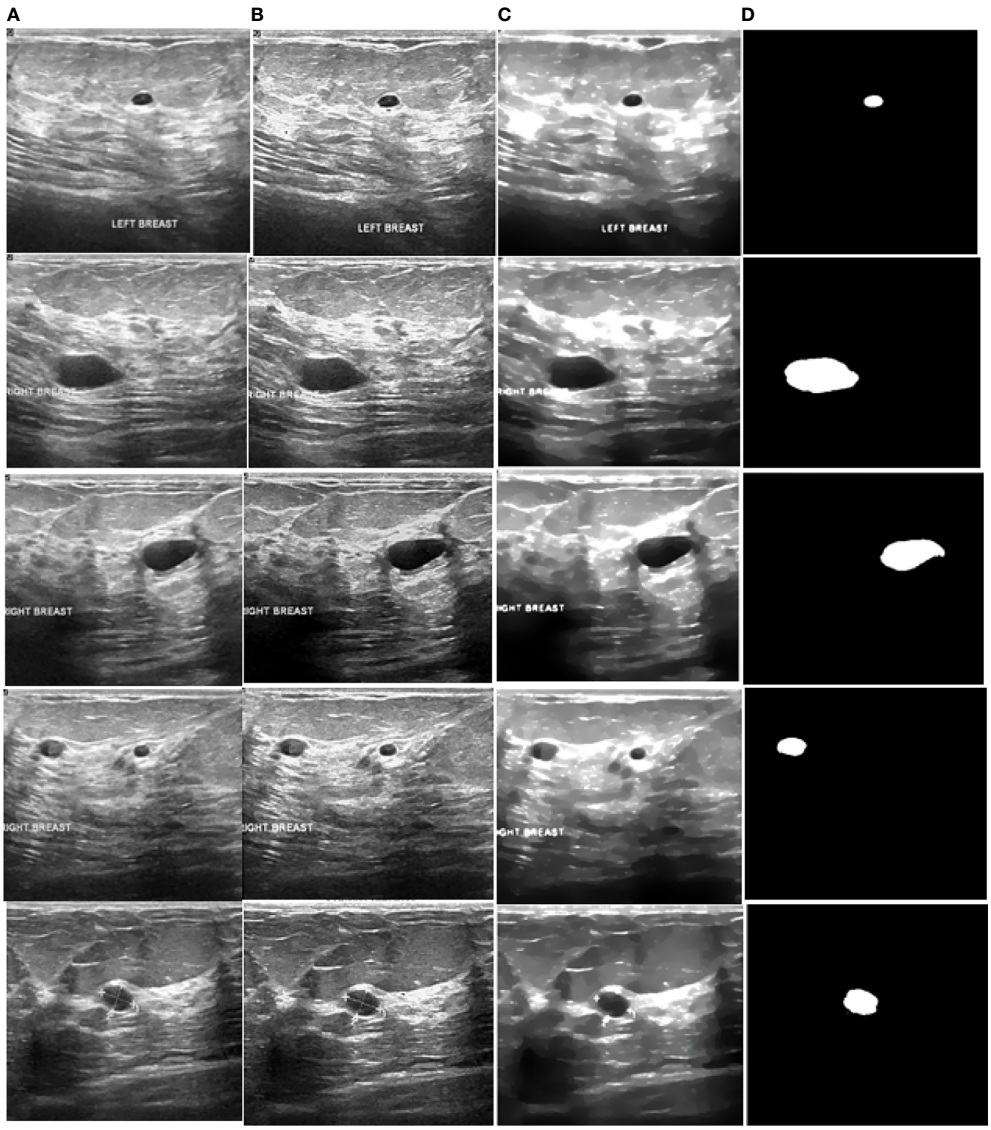


FIGURE 9 | Processed benign images. **(A)** US image, **(B)** enhancement, **(C)** despeckling, and **(D)** Segmentation.

This paper evaluated some widely used performance indicators such as the confusion matrix, receiver operating characteristic (ROC), classification accuracy, sensitivity, and specificity. Accuracy calculated the proportion between true positive and total data, and shall be mathematically represented in the following Equations 27, 28, and 29:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{27}$$

$$Specificity = \frac{TN}{TN + FP} \tag{28}$$

True positive is represented by TP, true negative is represented by TN, false negative is represented by FN, and false positive is represented by FP.

$$Sensitivity = \frac{TP}{TP + FN} \tag{29}$$

3.4 Comparison With the State-of-the-Art Existing Methods

In this paper, a CAD model in breast cancer diagnosis was proposed and implemented. The efficiency of the presented CAD system was analyzed with the current methods: SOM-SVM (15), LFOA-SVM (24), MBA-RF (31) and BAS-BPNN (32) in terms of statistical metrics and ROC curve.

Figure 12 and **Table 2** display the classification performance of the proposed GWO-tuned WNN with the existing classifiers SOM-SVM, LFOA-SVM, MBA-RF, and BAS-BPNN (15). integrated SOM and SVM [SOM-SVM] for developing a breast

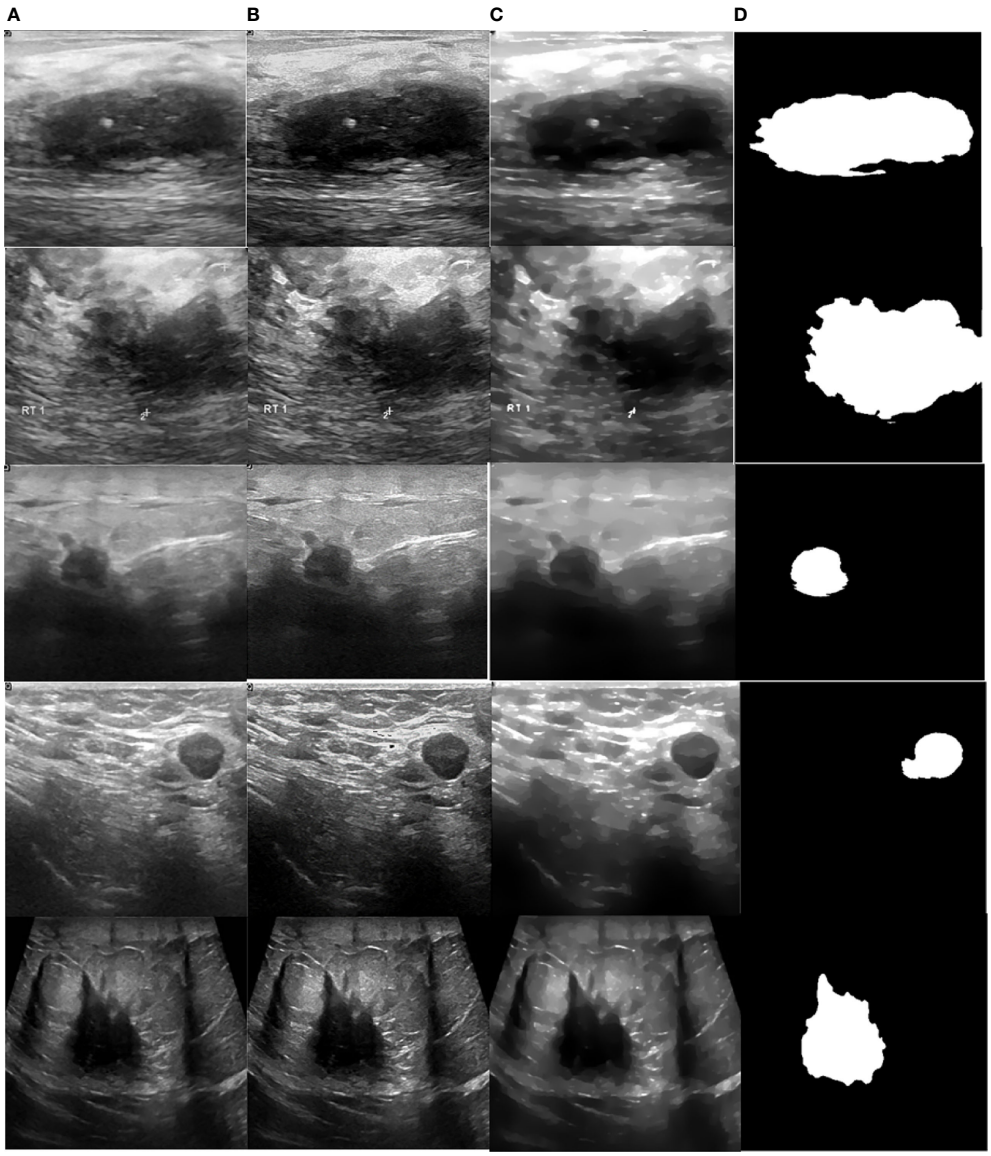


FIGURE 10 | Processed malignant images. **(A)** US image, **(B)** enhancement, **(C)** despeckling, and **(D)** Segmentation.

cancer diagnosis system. Here, the input images were preprocessed and divided into ROIs. The segmented ROIs and subregions were considered as bags and instances of bags, respectively. Subsequently, SOM was used for converting the instance space to concept space. Finally, SVM was incorporated for classifying the breast lesions. The major limitation here is that the method uses only local features for classification. Global features are also equally important for breast cancer diagnosis, thereby this is considered as a main drawback in this approach.

(24) implemented a framework by using SVM for breast cancer diagnosis. Here, the fruit fly optimization algorithm (FLA) modified by the Levy flight optimization algorithm (LFOA) was adopted for tuning the SVM parameters. High level features were obtained, normalized, and applied as input to the LFOA-SVM. However, this framework resulted in higher computational overhead, which was considered to be a focal drawback in this approach.

(31) examined the features of the modified bat algorithm (MBA) in selecting the feature vectors for breast cancer diagnosis. MBA employs the random sampling method to select features from the input data. Selected features are ranked and further employed to train the random forest (RF) classifier. This MBA-RF method exhibited comparatively better outcomes than the correlation-based feature selection method.

(32) formulated a strategy by integrating bio-inspired algorithms (BAs) and the back propagation neural network (BAS-BPNN). Here, the input data were subjected to preprocessing, feature extraction/selection, and classification.

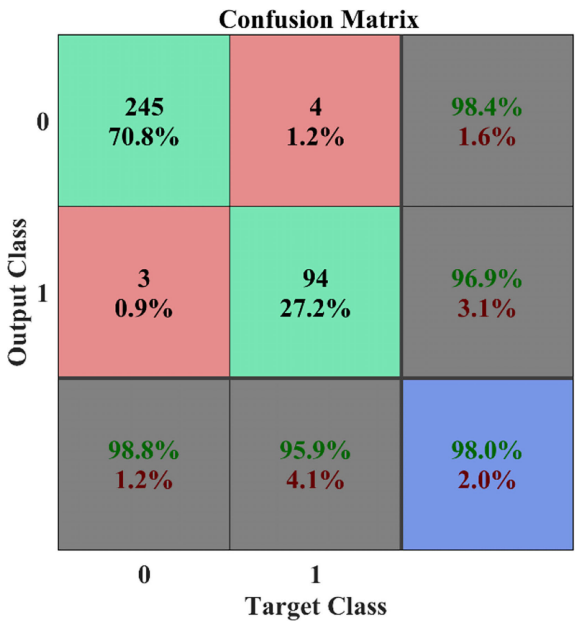


FIGURE 11 | Confusion matrix.

Moreover, three BAs were typically employed for selecting the features. Each algorithm selected corresponding subsets of features. The selected features were combined and classified by using BPNN.

(33) formulated three subnetworks, of which the first one contained the ROI, the second one contained the filtered ROI, and the third one was spread to the whole image. The first subnetwork in the CNN model was formulated using a single block, whereas the second subnetwork was formulated by two blocks and the third network was formulated by three blocks.

For datasets acquired using different scanners with different imaging protocols (34), used an efficient method by transfer learning to fine-tune the classification task in order to improve the accuracy. The mean accuracy reached 0.9 by the CNN hence accuracy in testing increased from 0.47 to 0.78 using CNN.

(35) suggested that the mixture of –spatial spectral features boost the classification precision of the CNN model which needs considerably fewer training parameters comparable with the best known system.

(36) formulated a conventional binary SVM classifier which was trained with seven features such as color and texture attributes obtained from the Gaussian model super pixels of the images belonging to WCE.

Numerous methods have been developed to analyze the strength of meta-heuristic algorithms for selecting salient features to improve the classification rate. However, these methods have some shortcomings like increased training time, overfitting, and slow convergence. Very few researchers have used meta-heuristic algorithms for tuning the parameters by

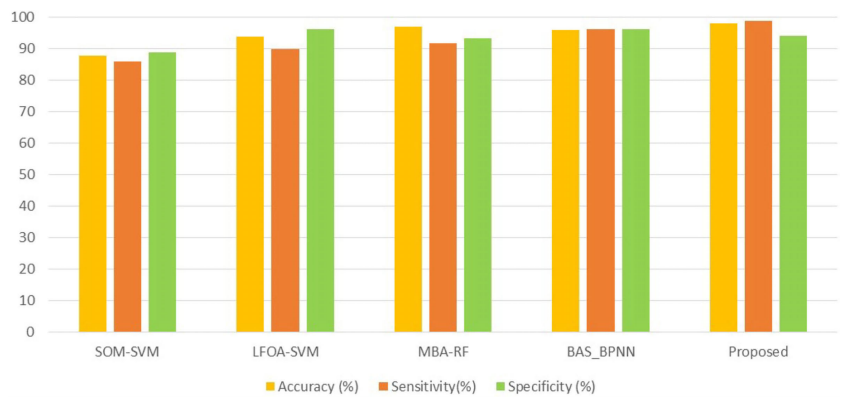


FIGURE 12 | Performance comparison with existing methods.

TABLE 2 | Classification performance of the proposed GWO-tuned WNN with existing classifiers.

Contributors	Data	Method	Accuracy(%)	Sensitivity(%)	Specificity(%)
(15)	US image	SOM-SVM	87.5	86.11	88.54
(24)	US images	LFO-ASVM	93.62	90	96.3
(31)	WDBC	MBA-RF	96.85	91.4	93.2
(32)	WDBC	BAS-BPNN	96.3	96.3	95.69
Proposed method	US images	GWO-WNN	98	98.8	95.9

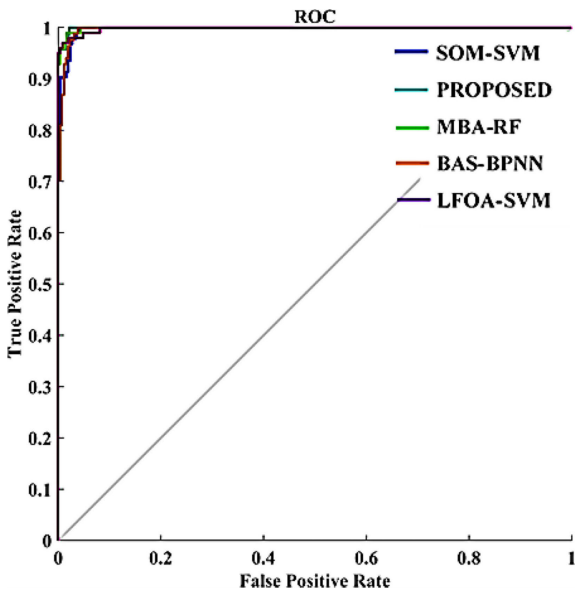


FIGURE 13 | ROC curve comparison.

machine learning. Therefore, we used the GWO algorithm for tuning the parameters of WNN to reduce the computational overhead and training time. In the proposed system, 10-cross fold validation is considered to test the efficacy.

Almost all previous research works employed BAS for feature selection. But, in this paper, we employed the GWO algorithm for optimizing the WNN parameters. Here, optimization aids to reduce computational cost and improve classification accuracy. As depicted in **Figure 13**, it was observed that existing methods were 87.5% (SOM-SVM), 93.62% (LFOA-SVM), 96.85% (MBA -RF), and 96.3% (BAS-BPNN) accurate in terms of classification (15, 24, 31, 32).

However, the proposed CAD system attained a classification precision of 98% which outperforms the other existing models SOM-SVM, LFOA-SVM, MBA-RF, and BAS-BPNN. Comparison outcomes indicate that the performance of the proposed GWO-WNN is better than other classifiers. This is because of the important scenario in the presented method of integrating GWO and WNN, and tuning WNN parameters with the GWO algorithm. GWO has a natural leadership mechanism which makes it more effective than other approaches. **Figure 13** provides the comparative analysis in terms of the ROC curve of the proposed scheme with other classifiers. From **Figure 13**, it can be seen that the area under the curve of the presented CAD system is greater than that of the state-of-the-art methods like SOM-SVM, LOFA-SVM, MBA-RF, and BAS-BPNN (0.99, 0.91, 0.93, 0.92, and 0.95). Therefore, this particular result highlights the effectiveness of the GWO-tuned WNN in US image classification. As a result, classification accuracy was improved with the proposed method.

4 CONCLUSION

Ultrasound imaging techniques are an important diagnostic tool for breast cancer diagnosis. The CAD model can assist physicians in generating more precise decisions. This paper proposed a CAD model for breast cancer diagnosis by the integration of a wavelet neural network and the grey wolf optimization algorithm. In this system, a sigmoid filter was used to increase the contrast of the images and IDAD to remove the speckle noise. An ROI was obtained from the preprocessed image, and subsequently the morphological and texture features were extracted and combined. The dimensions of the obtained features were reduced with PCA. Finally, GWO-WNN was executed to perform the classification task. Conventional WNN requires increased time for training and optimizing the parameters. This proposed scheme employed GWO for tuning the WNN parameters which eventually reduced the computation and training times. The simulation results clearly showed that the suggested CAD model system attained desirable classification accuracy when compared to its peers with respect to the classification accuracy. The leading advantages of the proposed GWO-WNN method include increased robustness, smaller training data, and faster convergence. Ultimately, since the GWO algorithm has a natural leadership mechanism, it makes it superior to other bio-inspired algorithms. Most of the former methods have employed a meta-heuristic algorithm for feature selection. But, the proposed CAD system used the meta-heuristic algorithm to tune the WNN parameters which helped to reduce the computational cost and avoid the local minima problem, thereby exhibiting improved performance. In future, few other meta-heuristic algorithms like cat swarm optimization, ant lion optimization etc., will be investigated to further improve the classification accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AA: Conceptualization, Methodology, Writing original Draft; PM: Funding Acquistion; MH: Funding Acquisition; SB: Supervision; SA: Resources. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Taif University Researchers Supporting Project (number TURSP-2020/26), Taif University, Taif, Saudi Arabia. KR: Supervision.

REFERENCES

1. Cengiz. *Cancer Facts & Figures* Vol. Vol. 5. . Atlanta, GA: American Cancer Society (2019).
2. Zarychta P, Badura P, Pietka E. Comparative Analysis of Selected Classifiers in Posterior Cruciate Ligaments Computer Aided Diagnosis. *Bull Pol Acad Sci: Tech Sci* (2017) 65:63–70. doi: 10.1515/bpasts-2017-0008
3. Kurek J, Świderski B, Osowski S, Kruk M, Barhoumi W. Deep Learning Versus Classical Neural Approach to Mammogram Recognition. *Bull Pol Acad Sci Tech Sci* (2018) 66:831–40. doi: 10.24425/bpas.2018.125930
4. Zhao Q, Qiu Y, Zhou G, Cichocki A. Comparative Study on the Classification Methods for Breast Cancer Diagnosis. *Bull Pol Acad Sci: Tech Sci* (2018) 66:841–8.
5. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ. A Metaheuristic Optimization Approach for Energy Efficiency in the Iot Networks. *Softw: Pract Exper* (2021) 51:2558–71. doi: 10.1002/spe.2797
6. Shabbir M, Shabbir A, Iwendi C, Javed AR, Rizwan M, Herencsar N, et al. Enhancing Security of Health Information Using Modular Encryption Standard in Mobile Cloud Computing. *IEEE Access* (2021) 9:8820–34. doi: 10.1109/ACCESS.2021.3049564
7. Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of Covid-19 Individuals Using Adaptive Neuro-Fuzzy Inference System. *Multimed Syst* (2021) 1–15. doi: 10.1007/s00530-021-00774-w
8. Dhanamjayulu C, Nizhal U, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of Malnutrition and Prediction of Bmi From Facial Images Using Real-Time Image Processing and Machine Learning. *IET Imag Process* (2021). doi: 10.1049/ipr2.12222
9. Poongodi M, Al-Shaikhli IF, Vijayakumar V. The Probabilistic Approach of Energy Utility and Reusability Model With Enhanced Security From the Compromised Nodes Through Wireless Energy Transfer in Wsn. *Int J Pur Appl Math* (2017) 116:233–50.
10. Poongodi M, Hamdi M, Malviya M, Sharma A, Dhiman G, Vimal S. Diagnosis and Combating Covid-19 Using Wearable Oura Smart Ring With Deep Learning Methods. *Pers Ubiquit Comput* (2021) 26:1–11. doi: 10.1007/s00779-021-01541-4
11. Pandithurai O, Poongodi M, Kumar SP, Krishnan CG. (2011). A Method to Support Multi-Tenant as a Service, in: *2011 Third International Conference on Advanced Computing*, pp. 157–62.
12. Chen HL, Yang B, Wang G. Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence. *J Med Syst* (2012) 36:2505–19. doi: 10.1007/s10916-011-9723-0
13. Poongodi M, Bose S. A Firegroup Mechanism to Provide Intrusion Detection and Prevention System Against Ddos Attack in Collaborative Clustered Networks. *Int J Inf Secur Priv (IJISP)* (2014) 8:1–18. doi: 10.4018/IJISP.2014040101
14. Poongodi M, Bose S. A Novel Intrusion Detection System Based on Trust Evaluation to Defend Against Ddos Attack in Manet. *Arab J Sci Eng* (2015) 40:3583–94. doi: 10.1007/s13369-015-1822-7
15. Ding J, Cheng HD, Huang J. Breast Ultrasound Image Classification Based on Multiple-Instance Learning. *J Digit Imaging* (2012) 25:620–7. doi: 10.1007/s10278-012-9499-x
16. Poongodi M, Bose S. (2013). Design of Intrusion Detection and Prevention System (Idps) Using Dgsotfc in Collaborative Protection Networks, in: *2013 Fifth International Conference on Advanced Computing (ICoAC)*. pp. 172–8.
17. Yang M. Robust Texture Analysis Using Multi-Resolution Gray-Scale Invariant Features for Breast Sonographic Tumor Diagnosis. *IEEE Trans Med Imaging* (2013) 32:2262–73. doi: 10.1109/TMI.2013.2279938
18. Ahila A, Stephen SA. Denoising of Ultrasound Breast Image Using Improved Oriented Speckle Reducing Anisotropic Diffusion Bilateral Filter (IADBF). *Rev T´ec Ing Univ Zulia* (2016) 9:35–40. doi: 10.21311/001.40.1.01
19. Poongodi M, Bose S, Ganeshkumar N. The Effective Intrusion Detection System Using Optimal Feature Selection Algorithm. *Int J Enterpr Netw Manage* (2015) 6:263–74. doi: 10.1504/IJENM.2015.073869
20. Flores WWG, Albuquerque CD, Pereira A, Catelli F. Infantosi. Improving Classification Performance of Breast Lesions on Ultrasonography. *Pattern Recogn* (2015) 48:1125–36. doi: 10.1016/j.patcog.2014.06.006
21. Chen Y, Huang Q. (2016). An Approach Based on Biclustering and Neural Network for Classification of Lesions in Breast Ultrasound, in: *2016 International Conference on Advanced Robotics and Mechatronics (ICARM)*, pp. 597–601.
22. Poongodi M, Manjula L, Pradeepkumar S, Umadevi M. Cancer Prediction Technique Using Fuzzy Logic. *Int J Curr Res* (2012) 4:106–10.
23. Jeyachandran A, Poongodi M. Securing Cloud Information With the Use of Bastion Algorithm to Enhance Confidentiality and Protection. *Int J Pur Appl Math* (2018) 118:223–5.
24. Huang H, Feng X, Zhou S. A New Fruit Fly Optimization Algorithm Enhanced Support Vector Machine for Diagnosis of Breast Cancer Based on High-Level Features. *BMC Bioinf* (2019) 20:676–87. doi: 10.1186/s12859-019-2771-z
25. Vijayalakshmi S, John A. Multi-Modal Prediction of Breast Cancer Using Particle Swarm Optimization With non-Dominating Sorting. *Int J Distribut Sensor Networks* (2020) 16:11. doi: 10.1177%2F1550147720971505
26. Adimoolam M, Karthi A, Govindharaju. A Hybrid Learning Approach for the Stage-Wise Classification and Prediction of Covid-19 X-Ray Images. *Expert Syst* (2021). doi: 10.1111/exsy.12884
27. Nikita J, Jhunthra. Prediction Modelling of Covid Using Machine Learning Methods From B-Cell Dataset. *Result Phys* (2021) 21:103813. doi: 10.1016/j.rinp.2021.103813
28. Gupta V, Jain N, Katariya P, Kumar A, Mohan S, Ahmadian A, et al. An Emotion Care Model Using Multimodal Textual Analysis on Covid-19. *Chaos Solitons Fractals* (2021) 144:110708. doi: 10.1016/j.chaos.2021.110708
29. Mirjalili S, Seyed M, Mirjalili A. Lewis. Grey Wolf Optimizer. *Adv Eng Softw* (2014) 69:46–61. doi: 10.1016/j.advengsoft.2013.12.007
30. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of Breast Ultrasound Images. *Data Brief* (2020) 28:104863. doi: 10.1016/j.dib.2019.104863
31. Jeyasingh S, Veluchamy M. Modified Bat Algorithm for Feature Selection With the Wisconsin Diagnosis Breast Cancer (WDBC) Dataset. *Asian Pacif J Cancer Prev* (2017) 18:1257–64. doi: 10.22034/APJCP.2017.18.5.1257
32. Christo VRE, Nehemiah HK, Minu B, Kannan A. Correlation-Based Ensemble Feature Selection Using Bioinspired Algorithms and Classification Using Backpropagation Neural Network. *Comput Math Methods Med* (2019) 2019:7398307. doi: 10.1155/2019/7398307
33. Merati M, Mahmoudi S, Chenine A, Chikh MA. A New Triplet Convolutional Neural Network for Classification of Lesions on Mammograms. *Rev D'Intell Artifi* (2019) 33:213–7. doi: 10.18280/ria.330307
34. Ha W, Vahedi Z. Automatic Breast Tumor Diagnosis in MRI Based on a Hybrid CNN and Feature-Based Method Using Improved Deer Hunting Optimization Algorithm. *Comput Intell Neurosci PMCID* (2021) 2021. doi: 10.1155/2021/5396327
35. Mewada HK, Patel AV, Hassaballah M, Alkinani MH, Mahant K. Spectral-Spatial Features Integrated Convolution Neural Network for Breast Cancer Classification. *Sensors* (2020) 20:7506633–7506633. doi: 10.3390/s20174747
36. Rathnamala S, Jenicka S. Automated Bleeding Detection in Wireless Capsule Endoscopy Images Based on Color Feature Extraction From Gaussian Mixture Model Superpixels. *Med Biol Eng Comput* (2021) 59:969–87. doi: 10.1007/s11517-021-02352-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 A, M, Bourouis, Band, Mosavi, Agrawal and Hamdi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fpubh.2022.902123

SperoPredictor: An Integrated Machine Learning and Molecular Docking-Based Drug Repurposing Framework With Use Case of COVID-19

Faheem Ahmed[1†], *Jae Wook Lee*[1,2†], *Anupama Samantasinghar*[1], *Young Su Kim*[2], *Kyung Hwan Kim*[1], *In Suk Kang*[1], *Fida Hussain Memon*[1], *Jong Hwan Lim*[1] *and Kyung Hyun Choi*[1,2*]

[1] *Department of Mechatronics Engineering, Jeju National University, Jeju, South Korea,* [2] *BioSpero, Inc., Jeju, South Korea*

The global spread of the SARS coronavirus 2 (SARS-CoV-2), its manifestation in human hosts as a contagious disease, and its variants have induced a pandemic resulting in the deaths of over 6,000,000 people. Extensive efforts have been devoted to drug research to cure and refrain the spread of COVID-19, but only one drug has received FDA approval yet. Traditional drug discovery is inefficient, costly, and unable to react to pandemic threats. Drug repurposing represents an effective strategy for drug discovery and reduces the time and cost compared to *de novo* drug discovery. In this study, a generic drug repurposing framework (SperoPredictor) has been developed which systematically integrates the various types of drugs and disease data and takes the advantage of machine learning (Random Forest, Tree Ensemble, and Gradient Boosted Trees) to repurpose potential drug candidates against any disease of interest. Drug and disease data for FDA-approved drugs ($n = 2,865$), containing four drug features and three disease features, were collected from chemical and biological databases and integrated with the form of drug-disease association tables. The resulting dataset was split into 70% for training, 15% for testing, and the remaining 15% for validation. The testing and validation accuracies of the models were 99.3% for Random Forest and 99.03% for Tree Ensemble. In practice, SperoPredictor identified 25 potential drug candidates against 6 human host-target proteomes identified from a systematic review of journals. Literature-based validation indicated 12 of 25 predicted drugs (48%) have been already used for COVID-19 followed by molecular docking and re-docking which indicated 4 of 13 drugs (30%) as potential candidates against COVID-19 to be pre-clinically and clinically validated. Finally, SperoPredictor results illustrated the ability of the platform to be rapidly deployed to repurpose the drugs as a rapid response to emergent situations (like COVID-19 and other pandemics).

Keywords: drug repurposing, COVID-19, machine learning, databases, data analytics, host proteomes, molecular docking

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Theodora Katsila,
National Hellenic Research Foundation, Greece
Praise Young,
University of Eastern Finland, Finland
Shahab S. Band,
National Yunlin University of Science and Technology, Taiwan
Jude Osamor,
Glasgow Caledonian University, United Kingdom

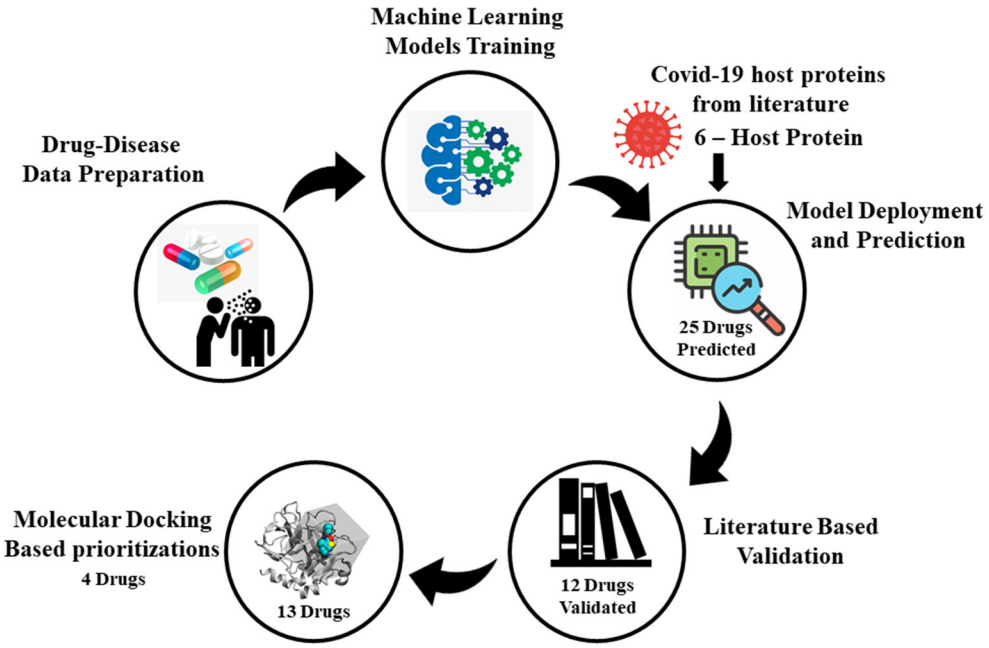
***Correspondence:**
Kyung Hyun Choi
amm@jejunu.ac.kr

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *22 March 2022*
Accepted: *02 May 2022*
Published: *16 June 2022*

Citation:
Ahmed F, Lee JW, Samantasinghar A, Kim YS, Kim KH, Kang IS, Memon FH, Lim JH and Choi KH (2022) SperoPredictor: An Integrated Machine Learning and Molecular Docking-Based Drug Repurposing Framework With Use Case of COVID-19. Front. Public Health 10:902123. doi: 10.3389/fpubh.2022.902123



GRAPHICAL ABSTRACT | Illustration of SperoPredictor repurposing framework. It starts with data collection followed by machine learning models training, model deployment, literature, and molecular docking-based validation.

INTRODUCTION

Coronavirus (CoV) normally targets the respiratory tract of humans (mammals), resulting in mild-to-severe respiratory tract infections (1). In the last two decades, severe acute respiratory syndrome (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) referred to as pathogenic human coronaviruses, have caused global epidemics with high mortality and morbidity rates (2) with a burden on the worldwide economy (3). In December 2019, Wuhan (China) experienced the third coronavirus pandemic, called novel coronavirus (SARS-CoV-2 or 2019-nCoV) disease 2019 (4). Over 79,000 confirmed cases with more than 2,600 deaths from COVID-19 or SARS-CoV-2 outbreak worldwide were reported as of 24 February 2020 along with contact transmission due to human-to-human interaction (5). Moreover, according to the World Health Organization (WHO), as of 15 February 2022, there have been a total number of 414.3 M confirmed cases of COVID-19 with 5.846 M deaths worldwide, while the United States, India, and Brazil remain the worst-hit countries (6), with worldwide economic costs of more than $16 trillion. In this connection, studies have been conducted for COVID-19 health and exploratory data analysis for classification, comparative analysis (7), and prediction using machine learning, such as done in (8) specifically targeting the Mexican and Brazilian patients. The results from the study consider data from patients under the age of 0–120 years demonstrating the application of big data technologies along with machine learning. Similarly, the relation between the spread rates of COVID-19 in high-risks countries is also studied in (9) where Pearson correlation was considered to first study the relationship between the countries under study followed by the use of PCA to categorize the countries based on the spread rates of COVID-19. Moreover, a vast number of efforts from various national and international research groups around the world have been made to obtain effective drugs for COVID-19. To date, the FDA has approved only one antiviral (oral) drug for the treatment of COVID-19, and it is regarded as a major step forward in the fight against COVID-19 (10). Owing to the lack of effective oral drugs for COVID-19, there is an urgent need to develop effective treatments for 2019-nCoV or SARS-CoV-2.

Additionally, witnessing the pandemics over the last two decades and the emergence of new diseases, the total money diverted to pharmaceutical and biomedical research has significantly increased, increasing the annual number of novel treatments approved by the US Food and Drug Administration (FDA) slightly (11). As estimated in a study, pharmaceutical companies invested $802 million in 2013 which increased to $2.6 billion in 2015 owing to the development of novel drug molecules approved by the FDA (12). Thus, the development of novel drug molecules until the approval of the FDA requires approximately $2–$3 billion investment and a time of approximately 12–15 years, with $< 10\%$ chance of success. This renders drug development a risky process that suffers from high attrition rates, substantial costs, and a slow pace (13). An effective alternative to *de novo* drug development is drug repurposing, the process of determining new indications of the already approved (FDA), failed/abandoned, or investigational drugs for use in other diseases (14). Drug repurposing reduces the time and money spent on *de novo* drug development and clinical trials (15–17) because the safety and efficacy data of the drugs to be repurposed are already known. In addition, sufficient investment of time and money is still required for drug repurposing when opting

for experimental approaches (18). The problem has been solved with the advent of computational approaches which have come along with enormous computing power to digest vast amounts of heterogeneous data for the testing of the systematic drug repurposing hypothesis (12, 13, 15–17). Moreover, targeting a single viral protein often results in high drug resistance owing to the rapid evolution of the virus genome (1). Conventional methods based on the three-dimensional structure of proteins and ligands are limited by the unavailability of protein structures that are normally encountered for most viral and human targets. Apart from that disease diagnosis studies are performed, such as done in (19). In the said study, significant features were ranked using a random tree, decision tree, support vector machine (SVM), and chi-squared decision tree models. The study results show the significance of the RT models which ultimately outperform the other models.

Moreover, *in silico* methods that predict drug–target interactions (DTI) or associations, use machine learning to train, test, and validate the machine-learning-based prediction models. DTIs are the pairs of drugs and their targets interacting with each other. Moreover, these trained prediction models are subsequently deployed to discover novel DTIs. The data on which feature-based machine-learning models are trained can be obtained from publicly available online databases or in-house laboratories. The well-known databases for DTIs and target–target interactions are DrugBank (20), KEGG (21), and STITCH (22–24). When discussing the representation of drugs and target data, people have presented them in many different forms; for example, drugs are represented by their chemical structures (25), drug expression profiles (26) side effects (27), and Anatomical Therapeutic Chemical (ATC) codes (28), while targets have been represented using their gene ontology information (28), genome sequence (21), protein–protein interaction (29), and disease associations (30). Moreover, many successful feature-based computational methods have been developed over the past decade. Feature-based methods are standard methods that require feature vectors of a fixed length as input (31). According to the literature, an earlier feature-based method (32) was used for target representation; this method performs binary vector representation of the drugs, indicating the presence or absence of functional groups in the chemical structure of the drugs with amino acid composition. Simply put, if a drug is represented by d, then the feature vector of the drug would be $[d_1, d_2, d_3,..., d_d]$. Similarly, for target t, it would be $[t_1, t_2, t_3,..., t_t]$. The drug and target feature vectors are subsequently concatenated to form a drug–target pair (d,t). In addition, many feature-based methods have been presented based on random forest (RF) (29, 30), rotation forest (33), and extremely randomized trees (34) that are formed by the ensemble of decision trees. In the given studies, many of the important features are not considered. Specifically, in most studies, only two or three features are focused on, missing the other important information. While in our work, unlike other studies, SperoPredictor® uses multiple drugs (four drugs) and disease (three diseases) features unified in binary strings covering diverse aspects of drugs and diseases data. The use of various features in our study make the classification models more generalized and accurate on unseen data. Additionally, other feature-based methods include relevance vector machines (35), KNN (fuzzy) (36), and deep learning (37–39). Additionally, most of the studies are focusing on single ML model trained and then deployed, while in our study we have trained multiple machine-learning models which are later stacked together to synergize the prediction confidence. Once the predictions are taken from the deployed machine-learning models, the need for further validation of the drugs is highly needed which most of the studies are lacking. Here, a bifold molecular docking validation is performed for the predicted drugs with host COVID-19 proteins. Moreover, there is dire need of well-trained models that are fast, accurate, and ready to be deployed as rapid response, such as SperoPredictor. Further it can also be deployed to predict the drugs for a given disease as input to the deployed model and can find the new indication for drugs given as input.

In this study, we present an integrated machine learning and molecular docking-based drug repurposing framework that systematically uncovers undiscovered DTIs. This approach is based on the notion that (i) given known and unknown sets of DTI data enriched with various aspects of drug and disease features (**Figure 1A**), (ii) machine-learning models can be trained (**Figure 1B**). The trained models (iii) can be deployed to determine new DTI pairs for a given set of proteins functionally associated with diseases causing viral infection, such as COVID-19 (**Figure 1C**). We employed this approach along with (iv) the confirmation of the indications from the literature survey and (v) by performing molecular docking to further ensure the interactions (binding affinity) between predicted drug–target pairs as shown in **Figure 1C**. Following this approach, 25 potential drug candidates were predicted for COVID-19. Literature-based validation confirmed that 12 out of 25 (48%) drug compounds were already used in COVID-19. For the remaining 13 compounds, molecular docking was performed. Based on the docking results and prediction scores, four potential drugs were suggested to be pre-clinically and clinically validated for COVID-19.

MATERIALS AND METHODS

In this step, we followed the five steps shown in **Figure 1** to identify the repurposed drugs for COVID-19 (drug repurposing).

Dataset Preparation

The dataset contains four drug features (drug chemical structures, side effects, target sequences, and drug-targeted genes) and three disease features (disease gene sequences, disease specificity indices, and disease observable traits).

Chemical Structures

Chemical structures (40) in SMILES form have been used in many studies. Simplified molecular-input line-entry system (SMILES) is the form of a line notation for describing the structures of chemical species and SMILES was obtained from DrugBank (41), PubChem (42), and ChEMBL (43) during 2021–2022 using application programming interface (API) in KNIME as shown in **Supplementary Figure 1**. To make the data machine-learning ready, one-hot encoding (44, 45) was done to convert SMILES data into a 1,430-dimensional vector called the Drug-S Vector. The vectors were subsequently collected

as the bit vector using the KNIME node called "Create Collection Column."

Drug–Target Sequences and Genes

Each drug targets one or more proteins (18, 46) and receptors on proteins are the most prominent targets for drugs. In this study, the drug–target sequences were extracted from Uniprot and DrugBank during the academic year 2021–2022 (47). Initially, DrugBank IDs were mapped to Uniprot IDs followed by using API in KNIME to retrieve sequence data from Uniprot. The "Sequence Properties" node in KNIME is used to extract the properties, such as the number of positively and negatively charged residues, mol-weight, hydrophobicity, and aliphatic index. Finally, one-hot-encoding

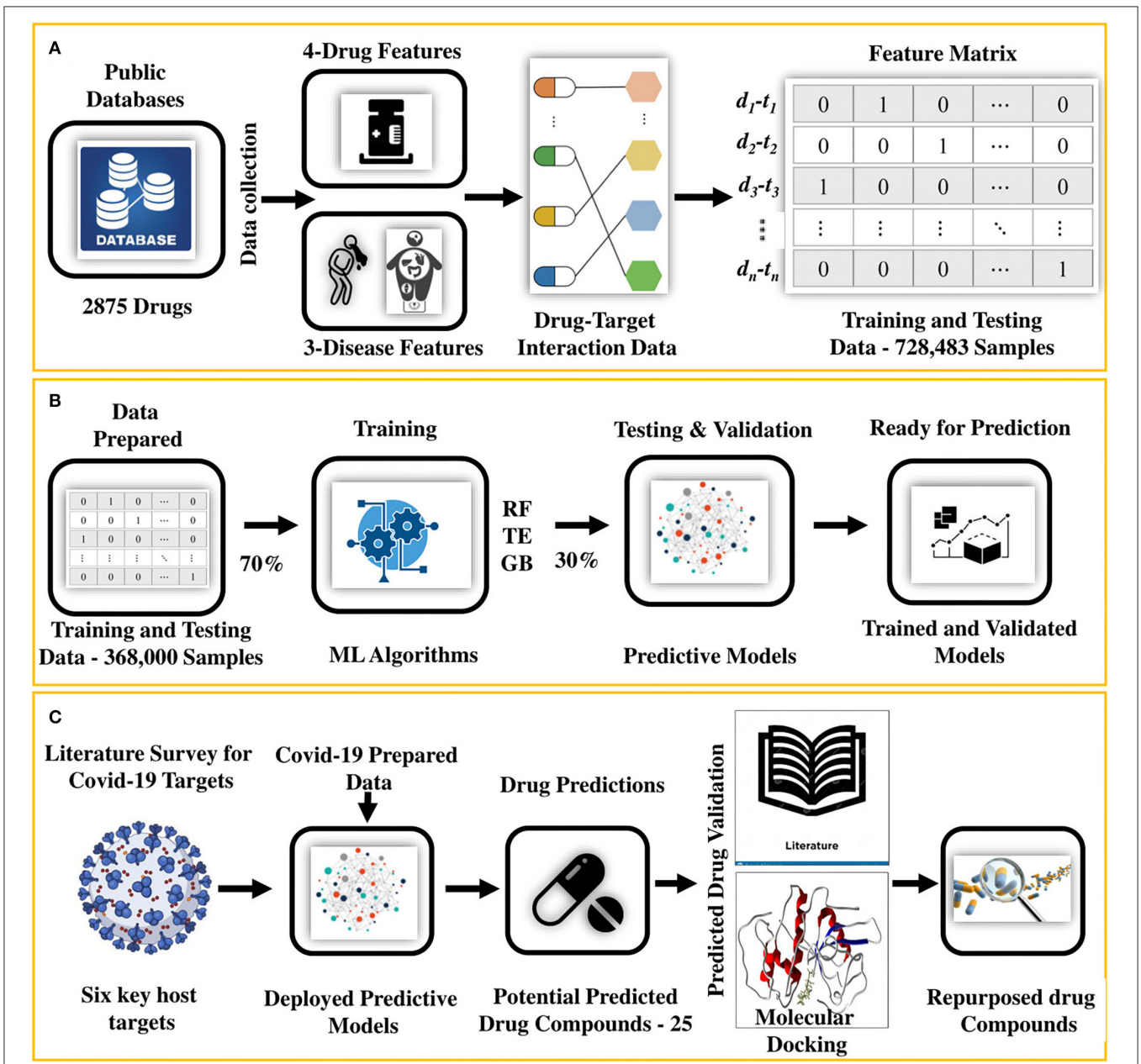


FIGURE 1 | An integrated machine learning and molecular docking-based ensemble approach for drug repurposing. **(A)** It shows that the drug and disease data collected from the public databases are first arranged and then transformed into a feature matrix, **(B)** followed by training and testing of the predictive models (RF and TE). **(C)** After training and testing, the data for the COVID-19 host targets are prepared from the literature review and fed to the predictive model for getting the potential repurposed drugs. Potential hits are then validated by using literature mining and collecting the evidence in the form of papers, patents, and database evidence. For the remaining predictions, molecular docking is performed to find the binding affinity of the drugs. Finally, the drugs with the highest binding affinity are prioritized as the potential repurposed candidates ready for the preclinical and clinical tests.

(48) was applied to extracted features and drug genes to transform features into the bit representation of the 2,294-dimensional vector.

Drug Side Effects

As a common hypothesis, drugs with similar side-effect profiles share similar therapeutic effects through a shared mechanism of action (49). The drug side-effect data of 1,430 FDA-approved drugs were downloaded from the SIDER database (27). SIDER provides mapping to PubChem which was used to map side effects to drug structures, protein sequences, and gene data, as shown in **Figure 2**. To make side-effect data machine-learning-ready, one-hot encoding was used in the KNIME analytics platform as shown in **Supplementary Figure 2** resulting in a binary vector string of 5,868-dimension called Drug-Se.

Disease Gene Sequences and DSI

In recent years, there has been a tremendous increase in the knowledge accumulation of gene–disease associations. It is important to facilitate clinical practice using this knowledge (50). In our study, gene-disease associations (GDA) were collected from DisGeNET (51–53) that contains more than 400,000 GDA across all databases (54). Moreover, the Uniprot IDs were used along with gene symbols from DisGeNET to extract gene sequences from the Ensemble database (55, 56) using API in KNIME (**Supplementary Figure 1**). Drug sequences were converted using one-hot encoding into a binary string vector. Additionally, for the disease specificity index (DSI) (51), the vectors generated are referred to as Dis-Ge for one-hot-encoded gene sequences of 1,321 length. Whereas, the DSI values are left named as DSI.

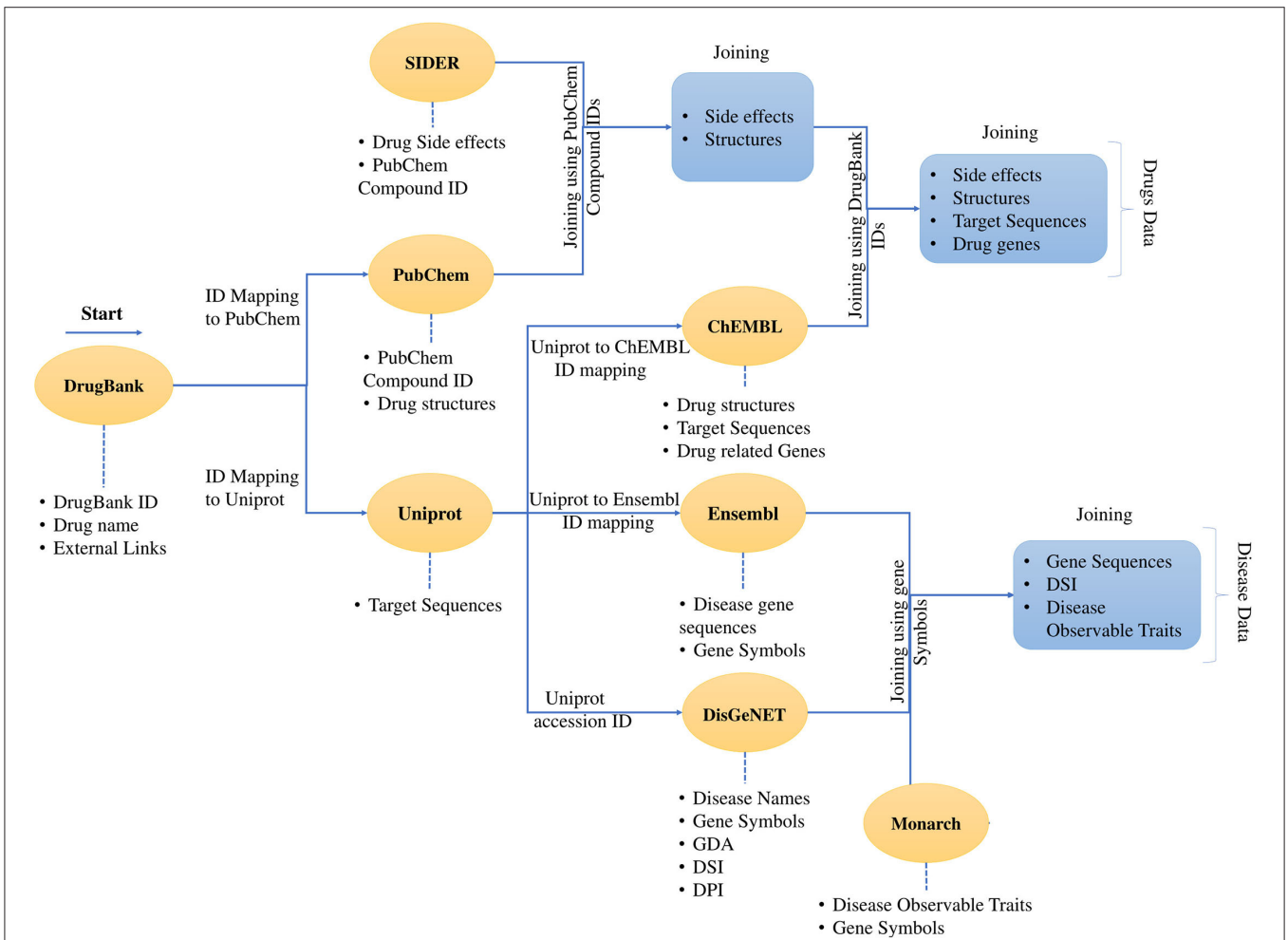


FIGURE 2 | Process of data collection and data mapping. It shows the creation of a dataset and merging of different data sources using cross-database identifiers. The dataset contains the data for the drugs and the data for the diseases. The data download was started with DrugBank data with external links. Then the data was mapped to PubChem and Uniprot. It was followed by the mapping of the SIDER data using the PubChem IDs and the mapping continues through ChEMBL and all the four drug features are presented in a tabular form with identifiers from mentioned databases. The same process follows for the disease data, and finally, the disease data is also represented in the tabular form. The developed dataset contains the drug and disease-related data enriched with various features.

TABLE 1 | Drug and disease features in the dataset created.

Dataset	Drug features				Disease features			Dataset features	
	Drug structures	Drug–target sequences	Drug related genes	Drug side effects	Disease gene sequences	DSI	Disease observable traits	+ve	-ve
	1,430	2,294	2,893	5,868	1,321	1,321	5,212	368,483	360,000

The table mentions the total number of each feature in the dataset. It contains four drug features and three disease features. The details of the different features and their mapping process are discussed in detail in Section Dataset Preparation.

Disease Observable Traits

Disease observable traits have been utilized for drug discovery and development (57–59). Drugs discovered based on phenotyping screens have surpassed the drug screened and developed through a molecular drug–target-based approach (60, 61). In this study, 10,881 human diseases with 8,662 phenotypes were retrieved from the Monarch database (62). The phenotypes were encoded using a one-hot encoder which produced vectors (Dis-Tr) of 5,212 lengths as shown in **Table 1**.

Positive and Negative Samples

After collecting the known drug- and disease-related features data (DTIs), we assumed them to be positive samples, and the unknown interactions achieved through the randomized shuffling of the positive samples were assumed to be negative samples (63) followed by upsampling of the negative samples to create a balanced dataset. The rationale behind this is that conventional methods of unknown interactions between targeted drugs as negative examples may result in bias because unknown interactions between targeted drugs may contain undetected interactions between the targeted drugs. This was overcome by finding the duplication between the positive and negative samples and it was ensured that no pair from positive samples matches exactly to negative samples.

Model Prediction Techniques

For bioinformatics research, machine learning plays a significant role in the filtering and comprehension of patterns in a given dataset (63). Our proposed study presents machine-learning models trained on various aspects of the drugs and disease data. The statistics for the datasets used in this study are provided in **Table 1**. The overall purpose of the workflow was to predict DTIs for COVID-19. The process can be divided into five steps. First, the models were trained using the training data, and second, the trained models were tested and cross-validated using the testing and cross-validation data, respectively. In this study, different machine-learning models were tested and evaluated. The ML models (algorithms) used in this work are Random Forest (RF), Tree Ensemble (TE), and Gradient Boosted Trees (GB).

Random Forest

It contains multiple individual decision trees that function as divisions. Every individual decision tree is bounded by predictions guided by the class, layer, and the soundest prediction results in the model. RF performs efficiently with large types of data elements from one decision tree, and its accuracy can be maintained even if the data contain missing values. In addition, the fair training time complexity expressed as $O(n*log(n)*d*k)$ and excellent space complexity of the RF models given as $O(depth\ of\ tree\ *k)$ makes it a good choice in such applications. In the given time and space complexity notations, "n" shows the number of training examples, "d" shows the data dimensions, and "k" shows the number of models. In our work, we used the KNIME node "Random Forest Learner" with the parameters shown in **Supplementary Table 1**. The model (RF) in this study functions in two ways. First, it was trained, tested, and validated; then, the results from the model were used individually and combined or averaged with other models. Finally, predictions for the COVID-19 DTI data were obtained after training and testing.

Tree Ensemble

Similarly, we used the KNIME node "Tree Ensemble Learner" to develop the TE model. It contains an ensemble of decision trees. Typically, each tree in ensemble learning is developed using different sets of rows or columns. Whereas, rows are called records and columns are called attributes. The key idea behind ensemble learning is that a group of weak learners combine to form a strong learner. Moreover, like RF, the time complexity of the TE algorithm is fair, and it is represented by $O(n*log(n)*d)$ and the space complexity of the TE is excellent. Similarly, in the given time and space complexity notations, "n" shows the number of training examples and "d" shows the data dimensions. In, this work, we used the "Tree Ensemble Learner" with the parameters listed in **Supplementary Table 1**. Like RF, TE learning algorithms also function in two ways. Additionally, the data prepared for COVID-19 were used to predict the DTIs to repurpose potential drugs for COVID-19.

Gradient Boosted Trees

Aiming at classification, it learns gradient boosted trees. It uses regression trees in a shallow form. Along with RF and TE, we used the "Gradient Boosted Trees Learner" node in KNIME for model development. The time and space complexity of the GB trees is also like the other two. Following the other two, GB was trained, tested, and validated, and the results from the model were combined and averaged with other models. After training and testing, predictions for the COVID-19 DTI data were obtained. The parameters used for the GB trees are also given in **Supplementary Table 1**.

Training and Testing Procedure

The prepared dataset was split into three portions: training, testing, and validation using the train-test split method. For each of the classifiers (Section Model Prediction Techniques), 70% of the data were used as training data that were randomly drawn and contained both positive and negative samples. The remaining 30% of the data were used for performance testing (15%) and validation (15%).

While selecting the best models, we used a part of the training data and tuned the hyperparameters. This procedure was followed by the testing of the trained models with 15% of the data split for the testing purpose. The process was repeated to fine-tune the hyperparameters. Once the testing accuracy was sufficiently high, the validation data were used to perform the 10-fold cross-validation of the models. During this initial procedure, we eliminated GB because it did not perform appropriately, compared to RF and TE. Therefore, we developed and trained the models; the list of parameters is given in **Supplementary Table 1**. The models were subsequently run on the validation datasets to observe the predictive performance using the evaluation parameters (Section Model Evaluation Techniques) that testifies to the effectiveness of the training performed. At the end of the training, the best-performing models as per the Mathew correlation coefficient (MCC) and other important parameters were selected. Finally, the models were either used in the combination (averaging the output of the models) or only the RF.

Model Evaluation Techniques

The different parameters used to evaluate the machine-learning model are discussed below.

Accuracy

This is the ability of a classifier to differentiate between positive and negative samples correctly. To determine the accuracy, true positive and true negative should be known across all cases. The accuracy can be measured using Equation (i).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{(i)}$$

Here, TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

Precision and Recall

Precision refers to the correct ratio of positive reactions, whereas recall shows the ratio of positive samples that have been predicted correctly. Precision can be calculated as per Equation (ii), and recall can be calculated according to Equation (iii).

$$Precision = \frac{TP}{TP + FP} \quad \text{(ii)}$$

$$Recall = \frac{TP}{TP + FN} \quad \text{(iii)}$$

F1 Score

This measures the balance between precision (p) and recall in the system. Equation (iv) was used to calculate the precision.

$$F1\ Score = \frac{2*[\ Precision*Recall\ ]}{[\ Precision + Recall\ ]} \quad \text{(iv)}$$

Mathew Correlation Coefficient

The value of MCC ranges from −1 to 1. In this range, −1 is a zero-credible binary-learning method, whereas 1 is a completely confident binary-learning method. The calculation formula for MCC is given by Equation (v).

$$mcc = \frac{TP*TN - FP*FN}{\sqrt{((TP + FN)*(TN + FP)*(TP + FP)*(TN + FN))}} \quad \text{(v)}$$

Prediction Validation Techniques

After the training, testing, and validation, the models were deployed. The data for COVID-19 host targets were collected from the literature. The information sources of the collected host targets for COVID-19 are given in **Table 3** along with other information. The data were pre-processed, and the predictions were obtained from the deployed ML models. The predictions (new DTI) were validated in the following two ways.

Literature-Based Validation of the New DTI Predictions

After the new DTI predictions were obtained for the COVID-19 targets, we performed a manual literature search. We used combination models (RF and TE) and individuals, and for both types of models, literature-based evaluation was performed (64, 65). As a result, we labeled the new predicted DTIs valid when at least one publication or database evidence showed the use of the predicted drug in COVID-19. This can be further divided into two categories: first, the evidence from the publications that could mention either the computational approach or a combination of the computational approach along with wet-lab experiments, and second, the database evidence. The databases, such as DrugBank, mention drugs along with their indications and are updated periodically.

Molecular Docking

After initial validation based on the literature, molecular docking was performed using AutoDock Vina for the unconfirmed DTIs. Moreover, for docking experiments, the co-crystallized structures of the host COVID-19 targets were downloaded from the Protein Data Bank (RCPDB) along with their PDB IDs. PyMol (66, 67) visualization software was used to visualize the target protein and extract the binding sites after deleting the water molecules and removing the backbone ribbon structure of the protein. Additionally, all the ligand molecules were also removed. The prepared protein structures were later used as inputs to Auto Dock Vina (67) for molecular docking, where hydrogen atoms and Gasteiger charges were added to the structures.

For the ligands, 3D structures were downloaded from PubChem (68) in structured data format (SDF). The structures were first converted into Protein Databank (PDB) format in

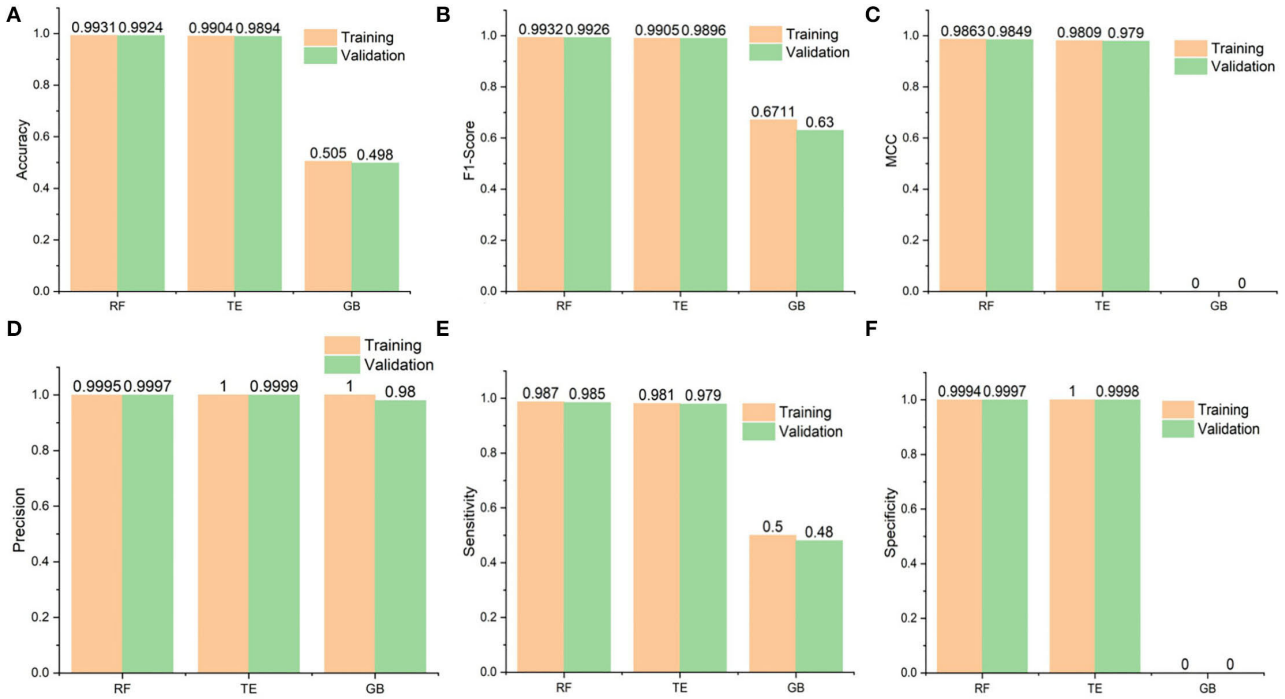


FIGURE 3 | Performance statistics of the algorithms used in the study. Additionally, it shows the training and 10-fold cross-validation results of the Random Forest (RF), Tree Ensemble (TE), and Gradient Boosted (GB) classifiers. **(A)** Shows the accuracy results of the classifiers and RF seems to perform best followed by the Tree Ensembl predictor, **(B)** in case of F1 scores, RF again performs better than the rest whereas the GB performance is worst, **(C)** whereas MCC values are also best for the RF but **(D)** Precision results are best for the TE during training and validation followed by the RF and GB, finally **(E,F)** for the Sensitivity and specificity, respectively, RF performs best and GB is constantly worst.

PyMol, followed by structure preparation in AutoDock Vina. The prepared ligand structures were named PDBQT format. Before performing docking, binding pockets were created using the known binding sites in the crystal structure, and docking was run for 10,000 iterations. The results for each ligand–target docking were collected separately and analyzed for binding affinity values. Finally, ligands with higher binding affinity values were prioritized, and then the re-docking procedure was carried out in AutoDock Vina with 10,000 iterations. Based on the docking and re-docking results, comparison potential COVID-19 drugs were suggested for preclinical and clinical validation.

RESULTS AND DISCUSSION

Here, we discuss the results obtained from our proposed machine-learning-based prediction models (SperoPredictor®) for DTI implementation on the generated drug–disease (DTI) dataset, followed by novel DTI prediction for COVID-19, their literature-based validation, and molecular docking-based validation.

SperoPredictor Training, Testing, and Validation

In our work, we handled the DTI prediction for COVID-19 as a binary classification problem. The developed BioSpero's SperoPredictor® accepts the drug–disease data in a one-hot-encoded format and performs binary predictions between drugs and targets as active (valid or interacting) or inactive (invalid or non-interacting) for a given set of protein targets. The predictions were further validated using literature-based evidence and molecular docking, as shown in **Figure 1**. Moreover, to train the models, we generated a multiple feature-based dataset for DTIs from multiple databases. The data collection for 1,430 drugs and 2,265 protein targets (**Table 1**) was partially performed through an application programming interface (API) in the KNIME analytics platform, and subsequently processed and transformed using KNIME nodes. API automatically extracts the data from databases using a set of identifiers and the data-mapping process is shown in **Figure 2**. To validate the models, we split the dataset into training (70%), testing (15%), and validation (15%) datasets. Moreover, 10-fold cross-validation was performed on the tested models, and the performance statistics are shown in **Figure 3**. The results are further summarized in **Supplementary Table 2**, illustrating the accuracy, MCC, F1 score, precision, sensitivity, and specificity, whereas the corresponding results are shown in **Figure 3**. The given parameters were calculated during testing and cross-validation.

The accuracies of the RF model during testing and validation were 0.9931 and 0.9924, respectively. Similarly, the TE accuracies remained at 0.9904 and 0.9894, respectively, during the testing and validation of the model. In addition, the precision-recall

values are also presented for the proposed models. It is important to mention that precision is also referred to as positive predictive value (PPV) and recall (sensitivity) is referred to as the true positive rate (TPR) (63); specificity, also called true negative rate (TNR), is also mentioned in **Supplementary Table 2**. Higher values for precision, recall, and specificity indicate the greater predictive performance of the trained models.

Comparison With the Existing Methods

In this section, we provide a comparison between SperoPredictor and three recent state-of-the-art methods. The first method proposed efficient machine-learning methods with a case study on COVID-19 (63). In these studies, SVM, RF, XGBOOST, and DBN were used. DTIs were predicted using RF and SVM for COVID-19. They predicted drugs for ACE2 with 100% confidence. Another method is based on the deep learning structure model (CNN) to gain the sequences (Amino acid) in 1D representation. According to the results, the use of CNN for data representation improved the performance, compared with traditional methods, and the performance among all models is quite visible. The other method offers a novel approach to developing negative DTI. In this study, many lasso models were used to combine multiple sets of features to examine the prediction power and DITs (69). Lasso DN's suggested comparing the performance to Lasso, support vector machine (SVM), standard logistic regression (SLG), and DNN models. Additionally, another study proposed DTI prediction using Lasso with RF. FP2 and PsePSSM fingerprints were used for feature extraction, followed by the removal of redundant information (70). Our method performs better than all the other methods, achieving the best results in testing and cross-validation, as shown in **Table 2**. As shown in **Figure 3** and **Table 2**, the highest accuracy achieved in our work is 0.9931 (99.31%) in the case of RF and 0.9904 (99.04%) in the case of the TE algorithm which is approximately 5% higher than the first (63), 6% higher than the second (69), and 7% higher than the third (70). Similarly, when the results are compared for other parameters, such as MCC, our proposed models exhibit better performance at 0.9863 for RF and 0.9809 for TE as compared to other studies done in (63, 69, 70). Another recent study done in (71) uses SVM and RF models to repurpose the drugs for COVID-19. The accuracies of the classifiers as shown in **Table 2** are 0.90 and 0.82, respectively, outperformed by RF and TE models presented in our study.

TABLE 2 | Model performance comparison between related and our work.

Study	The methods	Accuracy	MCC
Efficient machine-learning model (63)	RF	0.947	0.945
	SVM	0.93	0.917
Lasso-DNN method (69)	SVM	0.81	-
	ANN	0.9277	-
Lasso with random forest (70)	RF	0.9809	-
Repurposed drugs for COVID-19 using AI and ML (71)	RF	0.82	-
	SVM	0.90	-
Our proposed work	RF	0.9931	0.9863
	TE	0.9904	0.9809

The models from the related were listed along with their performance parameters. The accuracy and MCC of the Random Forest and Tree ensemble classifiers were matched with the state-of-the-art methods from the literature.

Application in COVID-19

De novo drug development takes 12–15 years with a $2–3 billion investment. Drug repurposing has emerged as an effective alternative solution to respond in case of epidemics and pandemics. In this regard, repurposing potential antiviral drugs for COVID-19 is the only solution to counter the sudden emergence of pandemics. Moreover, the recurrence of COVID-19 is significantly affected by the response of the human immune system. To date, there has been extensive research conducted involving different methods (such as the data-driven method) on COVID-19. These methods use different resources for data collection and analysis, such as DrugBank, TTD, STITCH, and ZINC. Additionally, antiviral drugs targeting COVID-19 can be placed in one of two categories. First, antiviral drugs target the host targets of the virus to impede the aggregation of the virus, and second, drugs boost the immune response over a wide spectrum (72).

Moreover, a study using the deep learning model (73) conducted pre-training on the interactions, and MT-DTI identified the EGFR receptor successfully as the drug–target that has been associated with 30 candidates out of 1,094 from the DrugBank. This study did not require 3D structural information to predict the interactions between drugs and targets. However, there is no evidence to support the positive action of drugs in COVID-19 (74). The drug Atazanavir exhibited reasonable efficacy and binding affinity toward COVID-19-target proteins. In addition, the FDA approved Remdesivir for use in patients aged 12 years and above (75). Another study performed drug repurposing for COVID-19 using a literature-based approach (76). They used a scientific approach called the literature-based discovery (LBD) and compared it with three other similar models. They concluded that semantic models are most suitable for drug repurposing.

A recent study (77) suggests the application of AI to accelerate the drug repurposing process (78). This further demonstrates the importance of different AI approaches and discusses the different drugs that are currently under clinical trials in various phases. Similarly, another study discusses how AI models are used in precision medicine and how AI models can accelerate drug repurposing in COVID-19. Furthermore, AI along with network-based approaches, can be a powerful and innovative alternative to drug repurposing. According to these reviews, AI-based tools can be used to reposition drugs for other human-related diseases, focusing on COVID-19.

Here, we focused on predicting DTIs using SperoPredictor, the proposed machine learning and molecular docking repurposing framework that was trained on drug–disease-enriched data (the corresponding statistics are shown in **Table 1**). The data were collected from multiple sources and mapped to a single table using the cross-database identifiers and the mapping process is shown in **Figure 2**. In our study, to predict the repositioning drugs for COVID-19, we deployed our trained models with high

TABLE 3 | Key host-target proteins in COVID-19.

	Uniprot	Entry name	Protein name	PDB ID	References
1	O15393	TMPS2_HUMAN	Transmembrane protease serine 2	7MEQ	(79, 80)
2	P09958	FURIN_HUMAN	Furin	5MIM	(81)
3	Q9BYF1	ACE2_HUMAN	Angiotensin-converting enzyme 2	7V8V	(82)
4	Q2M2I8	AAK1_HUMAN	AP2-associated protein kinase 1	5L4Q	(83)
5	O14976	GAK_HUMAN	Cyclin-G-associated kinase	4O38	(83)
6	P07711	CATL1_HUMAN	Procathepsin L	1CS8	(84)

The Uniprot identifiers for each target are mentioned along with their entry names and protein names. PDB IDs are also provided which are later used in molecular docking also. Additionally, references for each protein target are given.

TABLE 4 | Drugs predicted for COVID-19 with their prediction probability and reference links.

	DrugBank ID	Drug name	Targets	Prediction Confidence	References
1	DB01054	Nitrendipine	AAK1	0.909	(85)
2	DB12610	Ebselen	FURIN AAK1	0.916	(86, 87)
3	DB04954	Tecadenoson	AAK1	0.912	(88)
4	DB12831	Gabexate	ACE2 CTSL	0.912	(89)
5	DB12945	Dihydralazine	TMPRSS2 CTSL ACE2	0.946	(90)
6	DB13014	Hypericin	AAK1	0.901	(91)
7	DB13025	Tiapride	FURIN AAK1	0.917	(92)
8	DB13132	Artemisinin	CTSL ACE2	0.964	(93, 94)
9	DB13141	Ambroxol acefyllinate	TMPRSS2 CTSL ACE2	0.94	(95, 96)
10	DB13620	Potassium gluconate	AAK1	0.911	(97, 98)
11	DB13875	Harmaline	GAK FURIN ACE2	0.943	(99, 100)
12	DB13876	Brofaromine	TMPRSS2 CTSL ACE2	0.948	(101)

The mentioned drugs are the results of the drugs from the Random Forest model and model combination (Random Forest + Tree Ensembl). The first two drugs are the result of the combination models. Moreover, the table contains the drugs and their DrugBank IDs which are validated from the literature for their use in COVID-19. It supports and adds to the credibility of our developed ML models. Additionally, the COVID-19 targets in the form of Uniprot IDs along with prediction confidence and literature-based evidence.

accuracy and specificity of 99.3–99.94% (RF) and 99.0–99.98% (TE), respectively.

To get the predictions for potential repurposed drugs for COVID-19, a literature survey was conducted, and the key host-target proteins of COVID-19 were collected. The target proteins are shown in **Table 3**. After collecting the targets, the data, including the observable traits, Ensembl gene ID, and DSI index, for these targets were prepared, as shown in **Supplementary Table 5**. The observable traits were collected from the Monarch database, whereas gene IDs and DSI indices were collected from the Ensemble database. The data were subsequently fed into the workflow for processing and gene sequence extraction, and the workflow is shown in **Supplementary Figure 3**. Following the processing and transformation of the COVID-19 target data, they were provided as input to the deployed RF and TE model for the predictions, as shown in **Supplementary Figure 4**. The models were used in two configurations: first, the combination of both models was used, and as a result, two drugs were predicted; second, the RF model was used, and a total of 25 drugs were predicted.

Our strategy for validating the prediction was based on literature-supported evidence and molecular docking. Two drugs predicted from combination models (RF and TE) were validated from the literature, and evidence suggested their repurposed use in COVID-19. Whereas 12 out of 25 drugs predicted using RF were found in the literature, suggesting their use in COVID-19. This means 48% of the predicted drugs were known to have anti-COVID-19 activity. All predicted drugs, along with evidence found in the literature, are specified in **Table 4** with prediction confidence and literature evidence. Whereas, all 25 predicted drugs, along with the information, are specified in **Supplementary Table 3**.

Molecular Docking and Re-docking for Binding Affinity Prediction of Compounds Against SARS-CoV-2 Host Targets

AutoDock Vina was used to perform molecular docking between the predicted ligands (for SARS-CoV-2) and SARS-CoV-2 host protein targets. All the selected host COVID-19 targets are listed in **Table 3**. The binding sites and PDB IDs of the selected proteins are shown in **Supplementary Table 4**. Ligand structures were downloaded in 3D SDF format from PubChem and converted to the PDB format using PyMol software. The preparation of 13 ligands (not confirmed in the literature) and six targets was conducted in the AutoDock Vina environment. Each ligand molecule was docked against all the six selected host COVID-19 protein targets. The binding affinity results are shown in **Figure 4**. As a result of molecular docking, the lowest binding affinity score was found as −5.4 (kcal/mol), and the highest score was −10 (kcal/mol). The top-scoring drugs for Adaptor Protein 2-Associated Kinase 1 (AAK1) with binding affinities ranging from −9.0 to −10.0 kcal/mol were Balaglitzone ($\Delta G = -9.9$ kcal/mol), Cortivazol ($\Delta G = -9.9$ kcal/mol), Ganaxolone ($\Delta G = -9.1$ kcal/mol), and Velusetrag ($\Delta G = -9.1$ kcal/mol) (**Supplementary Table 6**). Similarly, for Cyclin-G-Associated Kinase (GAK), the highest-scoring drugs were Balaglitzone ($\Delta G = -9.6$ kcal/mol), Velusetrag ($\Delta G = -10.0$ kcal/mol), 16-alpha Bromoepiandrosterone ($\Delta G = -9.0$ kcal/mol), and Rolofylline ($\Delta G = -9.1$ kcal/mol) (**Supplementary Table 6**). For the Angiotensin-Converting Enzyme 2 (ACE2), Cortivazol ($\Delta G = -10.0$ kcal/mol)

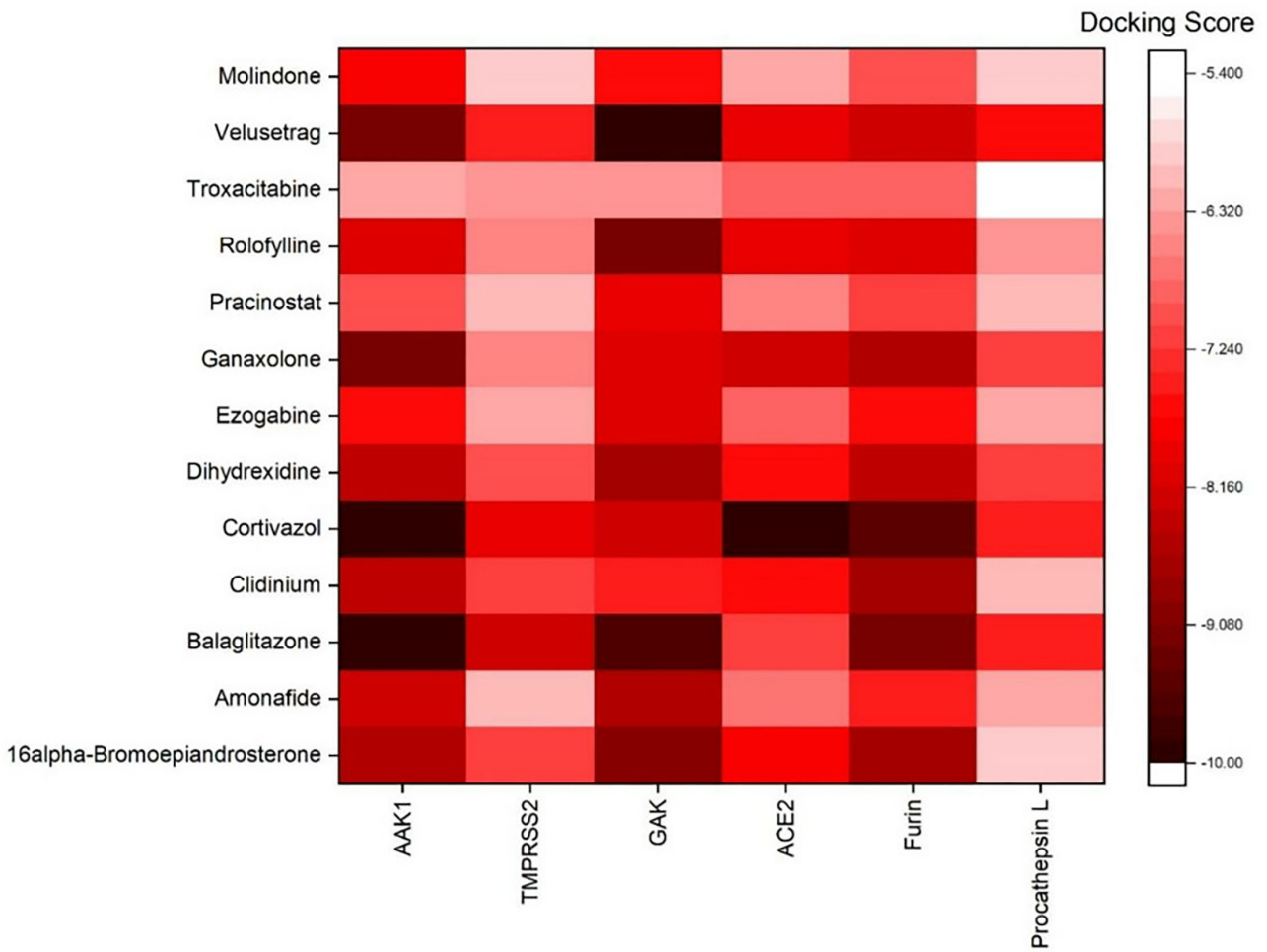


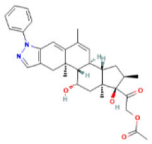
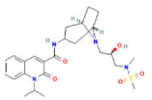
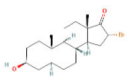
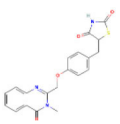
FIGURE 4 | Binding affinity or docking scores of the predicted repurposed drugs against six COVID-19 host proteins. The candidate predicted drugs are mentioned along the y-axis whereas the host COVID-19 protein targets are mentioned along the x-axis. The Docking score bar is showing the maximum to the minimum value, where the dark red color presents the maximum score of −10 and the white fed color shows −5.4. The highest binding affinities of 9.0 to −10.0 kcal/mol were achieved for Velusetrag-GAK, Cortivazol-ACE2, Balaglitazone-AAK, and Cortivazol-ACE2 complexes.

performed the best and for Furin, Cortivazol ($\Delta G = -9.4$ kcal/mol), Balaglitzone ($\Delta G = -9.1$ kcal/mol), and 16-alpha Bromoepiandrosterone ($\Delta G = -8.7$ kcal/mol) demonstrated the best results (**Supplementary Table 6**). Additionally, for the Transmembrane Protease Serine 2 (TMPRSS2) and Procathepsin L, the docking score of all the drugs was found to be below −8.0 kcal/mol, as shown in **Figure 4**. To further validate the top hit (six) drugs, re-docking was performed and results were found consistent as shown in **Supplementary Table 6**. Finally, the top hit predicted drugs interacting with high-binding affinity to at least two host targets based on the molecular docking score were analyzed and suggested as an effective treatment to be pre-clinically and clinically validated for COVID-19 (**Table 5**).

Moreover, the molecular interaction of the top four hit molecules (four drugs) interacting with at least two host COVID-19 targets based on the molecular docking scores are shown in **Figure 5**. This indicates the binding interaction along with the different host protein targets, their residues, and interaction scores highlighted in **Figure 5**. Based on the docking scores, small-molecule drugs (**Table 6**), such as Balaglitazone, are suggested for preclinical validation against Adaptor Protein 2-Associated Kinase 1 (AAK1) (**Figure 5A**), Cyclin-G-Associated Kinase (GAK) (**Figure 5C**), and Furin (**Figure 5H**). Similarly, the other top-ranked candidate drug Cortivazol is recommended against host COVID-19 targets, such as Adaptor Protein 2-Associated Kinase 1 (AAK1), Furin, and Angiotensin-Converting Enzyme 2 (ACE2), followed by Velusetrag, recommended for Cyclin-G-Associated Kinase (GAK) (**Figure 5D**), and Adaptor Protein 2-Associated Kinase 1 (AAK1) (**Figure 5I**); 16 alpha Bromoepiandrosterone is suggested for Furin (**Figure 5G**) as well as for Cyclin-G-Associated Kinase (GAK).

Furthermore, looking at the severity and state of emergency COVID-19 has caused, there is a dire need for effective

TABLE 5 | Finally suggested four drug candidates for COVID-19 with their target information.

	Drug bank ID	Drug name	COVID-19 targets	Free energy of binding (Kcal/mol)	Free energy of binding (Kcal/mol) Re-docking	Prediction score
1	DB13003	Cortivazol	Adaptor protein 2 associated kinase 1 (AAK1)	−9.9	−10.0	0.94
			Angiotensin-converting enzyme 2 (ACE2)	−10	−10.0	
			Furin	−9.4	−9.4	
2	DB12702	Velusetrag	Adaptor protein 2 associated kinase 1 (AAK1)	−9.1	−8.6	0.944
			Cyclin-G-associated kinase (GAK)	−10	−10	
3	DB05107	16-alpha Bromoepiandrosterone	Cyclin-G-associated kinase (GAK)	−9.0	−9.0	0.902
			Furin	−8.7	−8.7	
4	DB12781	Balaglitazone	Adaptor protein 2 associated kinase 1 (AAK1)	−9.9	−10.1	0.946
			Cyclin-G-associated kinase (GAK)	−9.6	−9.7	
			Furin	−9.1	−9.1	

Docking and pre-docking scores in kcal/mol are given which confirm the docking results as valid. Additionally, the prediction results of the drugs from SperoPredictor are also shown here.

treatments. One way to respond to the urgent need is to use Biotechnology platforms to swiftly repurpose the potential drug candidates. The emergence and re-emergence of pandemics emphasize the need of building accurate and robust drug repurposing platforms like SperoPredictor to make the process of drug discovery smooth and faster. Moreover, the proposed work offers many benefits, such as it contains ML models trained on drug–disease data enriched with multiple features. Drug data were enriched with four features, however, disease data were enriched with three features. Rigorously trained ML models were tested, cross-validated, and can be applied to any disease, such as was deployed for COVID-19 to repurpose the drugs with high accuracy. Despite the highly accurate and confident prediction results limitations, such as lack of data (1,430 Drugs) and preclinical validation models, are acknowledged and will be taken as a potential future direction that will enhance the effectiveness of the SperoPredictor. Since the trained machine-learning models yield high testing and validation accuracies (RF = 99.3 % and TE = 99.04 %) and the predictions were confirmed from the literature (12/25–48%), these drugs cannot be directly used in clinical trials. Currently, this study excludes the preclinical and clinical validation of the prioritized drugs. Most importantly, the limitation of the available data for drugs and disease is by far the major limitation of this study followed by the lack of negative data samples (which were later upsampled) which could lead to machine-learning models resulting in a high false-positive rate. Additionally, the validation of the remaining (not confirmed from literature) DTI was performed using the molecular docking approach. A lot of research work has been done and molecular docking approaches are used for the prioritization of the anti-COVID-19 drug compounds, still these approaches are not perfect. To overcome the uncertainty and improve the confidence in the results, we used the docking in two steps: first docking of the remaining drugs (13) followed by the prioritization based on the docking scores as shown in **Supplementary Table 6** (6/13 were prioritized). Second, re-docking for the shortlisting of prioritized drugs was done and the results are shown in **Supplementary Table 6**. Finally, the four drugs were prioritized based on the number of COVID-19 targets and docking results. Each drug should have more than one target as shown in **Table 6**.

CONCLUSION

The past and present efforts are focused on accelerating the drug development process through an alternative approach called drug repurposing. In this study, we present SperoPredictor, a machine learning and molecular docking-based repurposing framework with a use case in DTI prediction for COVID-19. The RF (99.3%) and TE (99.03%) can classify the DTIs with high accuracy. SperoPredictor, is a generalized framework that can be deployed as a rapid response. Additionally, most previous methods have focused on the features of proteins and sequences, involving the physical and chemical properties

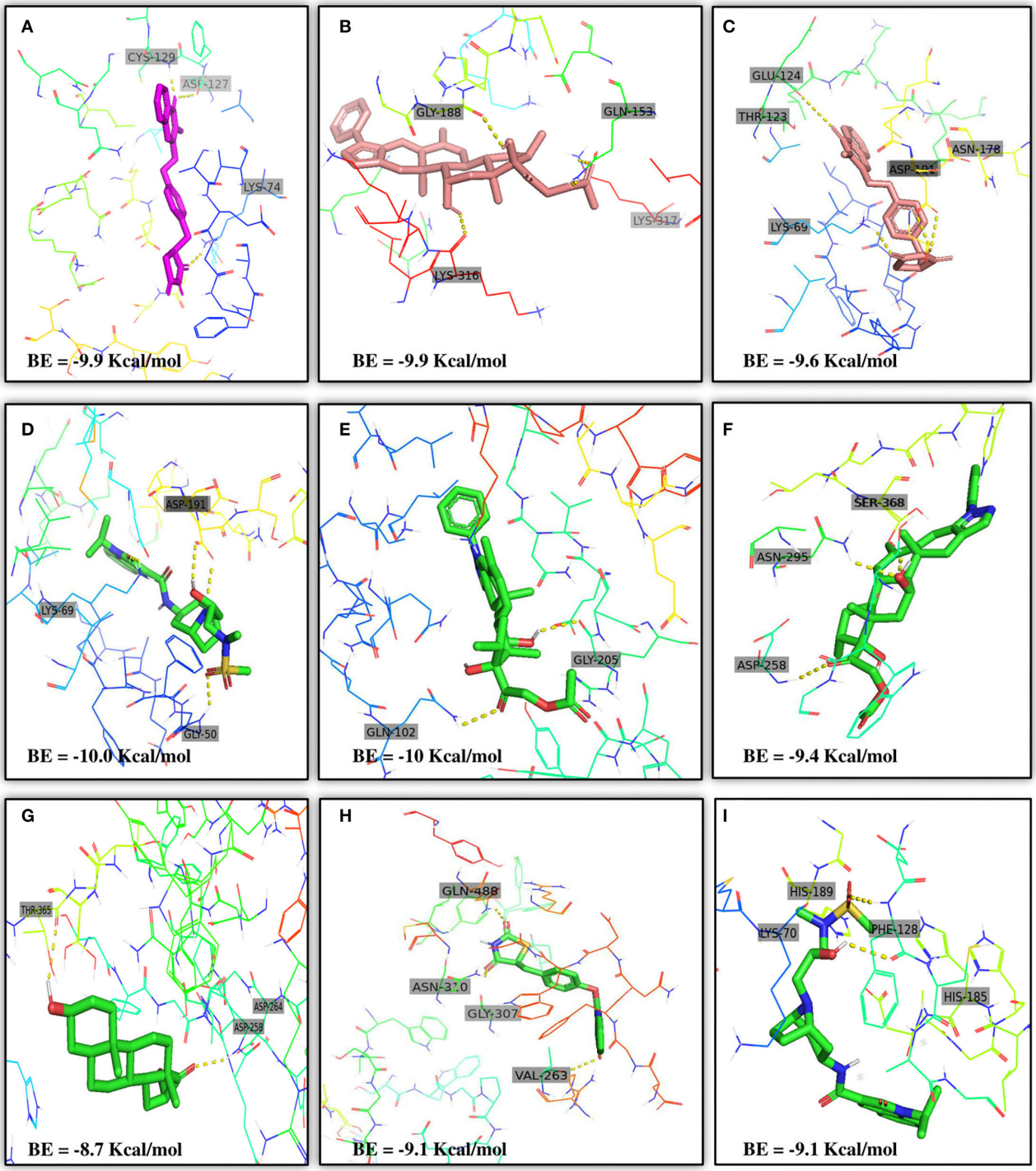


FIGURE 5 | Results of the molecular docking of the potential predicted drugs. Further, it shows the molecular interaction, residues, and scores. **(A,B)** It shows the interaction of Adaptor Protein 2-Associated Kinase 1 (AAK1) with Balaglitazone, Cortivazol, and Velusetrag, respectively. The docking score of the small molecules is −9.9 for the first two and −19.1 for Velusetrag which suggests strong binding affinity during the interaction. Additionally, in **(C,D)** interaction of Cyclin-G-associated kinase with Balaglitazone and Velusetrag is shown, respectively, with docking scores of −9.6 and −10.0. Whereas **(E)**, shows the molecular interaction between Angiotensin-converting enzyme 2 (ACE2) and Cortivazol with a docking score of −10 followed by **(F–H)** Cortivazol, 16 alpha-Bromoepiandrosterone, and Balaglitazone results with Furin. The docking score for Furin-Cortivazol complex is −9.4, for Furin-16 alpha-Bromoepiandrosterone, the docking score is −8.7, and for Furin-Balaglitazone, the docking score is −9.1. **(I)** Cortivazol binding affinity with Adaptor Protein 2-Associated Kinase 1 (AAK1).

TABLE 6 | Prioritized drugs with their original indication and host COVID-19 targets for which these drugs should be preclinically validated.

	Drug name	DrugBank ID	Original indication	Repurposed for	Host COVID-19 targets
1	Cortivazol	DB13003	Cluster headache-investigational (NCT00804895)	COVID-19	Adaptor protein 2 associated kinase 1 (AAK1) Angiotensin-converting enzyme 2 (ACE2) Furin
2	Velusetrag	DB12702	Gastroparesis and Alzheimer's-investigational	COVID-19	Adaptor protein 2 associated kinase 1 (AAK1) Cyclin-G-associated kinase (GAK)
3	16-alpha Bromoepiandrosterone	DB05107	Non-productive inflammation	COVID-19	Cyclin-G-associated kinase (GAK) Furin
4	Balaglitazone	DB12781	Diabetes mellitus, type 2–investigational	COVID-19	Adaptor protein 2 associated kinase 1 (AAK1) Cyclin-G-associated kinase (GAK) Furin

of drugs. The current proposed data integration method that accommodates the various aspects of drugs and diseases makes the predictions more confident. Similarly, a simple data pre-processing method saves running and processing time and space complexity. Moreover, using SperoPredictor models, we predicted 25 drugs (repurposed) for the SARS-CoV-19 host targets. According to the literature, 12 (48%) of the predicted drugs (25 drugs) have already been tested for SARS-CoV-19. Among the remaining 13 drugs based on molecular docking, re-docking and prediction confidence results of four drugs showing strong binding affinity are suggested for use in COVID-19. These drugs are Balaglitazone, Cortivazol, Velusetrag, and 16-alpha Bromoepiandrosterone. However, the majority of the predictions are validated by various literature sources and two-tier molecular docking validation is also performed. Thus, all the recommended drugs must be validated in various COVID-19 assays and clinical trials before being used in patients. We acknowledge the limitations in this study which will be taken as potential future directions and work along with possible drug combinations. This study is limited to the computational approach and excludes the *in vitro* validation. Additional limitations of this study include a limited number of drugs (1,430 drugs) and corresponding drug–target (2,294 targets) data along with the lack of negative data samples. Whereas, a higher number of drugs could contain and result in more potential anti-COVID-19 drugs and balanced data samples could result in better classification. Finally, the analysis of the novel DTI prediction for COVID-19 indicated that our approach could infer a list of novel DTIs that are practically applicable for drug repurposing. Moreover, the proposed approach can be used to repurpose the drug for any disease of interest following the same approach and this method can be applied in other ways to find the alternative uses for the existing drugs. However, in the future, the efficiency of this approach can be further enhanced by adding more data to train the models and using some state-of-the-art deep learning methods. In conclusion, predictive results of the SperoPredictor supported by literature evidence (12/25–48%) and good docking results between repurposed drugs (ligands) and SARS-CoV-19 host targets prove that our proposed models successfully identify potential repurposed candidates for COVID-19 treatment with high accuracy and confidence.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

FA contributed to methodology, conceptualization, formal analysis, investigation, validation, writing the original draft, molecular docking, and machine learning. JLe and AS contributed to methodology, formal analysis, investigation, and validation. YK, IK, and FM contributed to formal analysis, investigation, and validation. KK contributed to conceptualization, formal analysis, validation, and writing, reviewing, and editing the manuscript. JLi contributed to formal analysis and validation. KC contributed to conceptualization, investigation, resources, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program-Development of disease models based on 3D microenvironmental platform mimicking multiple organs and evaluation of drug efficacy) (20008413) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.902123/full#supplementary-material

REFERENCES

1. Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen KY. Coronaviruses-drug discovery and therapeutic options. *Nat Rev Drug Discov*. (2016) 15:327–47. doi: 10.1038/nrd.2015.37
2. Abramo JM, Reynolds A, Crisp GT, Weurlander M, Söderberg M, Scheja M, et al. Individuality in music performance. *Assess Eval High Educ*. (2012) 37:435. doi: 10.3389/fpsyg.2014.00661
3. De Wit E, Van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol*. (2016) 14:523–534. doi: 10.1038/nrmicro.2016.81
4. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. (2020) 395:507–13. doi: 10.1016/S0140-6736(20)30211-7
5. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. (2020) 382:1199–207. doi: 10.1056/NEJMoa2001316
6. *WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data*. Available online at: https://covid19.who.int/ (accessed January 05, 2022).
7. Shabbir A, Shabbir M, Javed AR, Rizwan M, Iwendi C, Chakraborty C. Exploratory data analysis, classification, comparative analysis, case severity detection, and internet of things in COVID-19 telemonitoring for smart hospitals. *J Exp Theor Artif Intell*. (2022) 1–28. doi: 10.1080/0952813X.2021.1960634
8. Iwendi C, Huescas CGY, Chakraborty C, Mohan S. COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients. *J Exp Theor Artif Intell*. (2022) 34-4:1—21. doi: 10.1080/0952813X.2022.2058097
9. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi A, Band SS. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. *Alexandria Eng J*. (2021) 60:457–64. doi: 10.1016/j.aej.2020.09.013
10. *Coronavirus (COVID-19) Update: FDA Authorizes First Oral Antiviral for Treatment of COVID-19 | FDA*. Available online at: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19 (accessed January 5, 2022).
11. Greene JA, Loscalzo J. Putting the patient back together — social medicine, network medicine, and the limits of reductionism. *N Engl J Med*. (2017) 377:2493–9. doi: 10.1056/NEJMms1706744
12. Avorn J. The $2.6 Billion Pill – Methodologic and Policy Considerations. *N Engl J Med*. (2015) 372:1877—9. doi: 10.1056/nejmp1500848
13. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*. (2018) 18:41–58. doi: 10.1038/nrd.2018.168
14. Pareek S, Huang Y, Nath A, Huang RS. The success story of drug repurposing in breast cancer. *Drug Repurposing Cancer Ther*. (2020) 173—90. doi: 10.1016/B978-0-12-819668-7.00006-3
15. Modiano P. Chapter 15. *Dora Bruder,* Vol. 1878. Oakland: University of California Press (2019). p. 62–66. doi: 10.1525/9780520962026-015
16. Cheng F, Hong H, Yang S, Wei Y. Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief Bioinform*. (2017) 18:682–97. doi: 10.1093/bib/bbw051
17. Cheng F, Murray JL, Rubin DH. Drug repurposing: new treatments for zika virus infection? *Trends Mol Med*. (2016) 22:919–21. doi: 10.1016/j.molmed.2016.09.006
18. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. (2016) 16:19–34. doi: 10.1038/nrd.2016.230
19. Joloudari JH, Hassannataj Joloudari E, Saadatfar H, Ghasemigol M, Razavi SM, Mosavi A, et al. Coronary artery disease diagnosis; ranking the significant features using a random trees model. *Int J Environ Res*. (2020) 17:731. doi: 10.3390/ijerph17030731
20. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*. (2013) 42:D1091–7. doi: 10.1093/nar/gkt1068
21. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. (2012) 40:D109–14. doi: 10.1093/nar/gkr988
22. Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen LJ, et al. STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Res*. (2013). 42:D401–7. doi: 10.1093/nar/gkt1207
23. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, Jensen LJ, et al. STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res*. (2009) 38:D552–6. doi: 10.1093/nar/gkp937
24. M. Kuhn, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res*. (2008) 36:D684–8. doi: 10.1093/nar/gkm795
25. Weininger D. SMILES, a chemical language and information system: 1: introduction to methodology and encoding rules. *J Chem Inf Comput Sci*. (1988) 28:31–36. doi: 10.1021/ci00057a005
26. Donner Y, Kazmierczak S, Fortney K. Drug repurposing using deep embeddings of gene expression profiles. *Mol Pharm*. (2018) 15:4314–25. doi: 10.1021/acs.molpharmaceut.8b00284
27. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res*. (2016) 44:D1075–9. doi: 10.1093/nar/gkv1075
28. Bergman U. The history of the Drug Utilization Research Group in Europe. *Pharmacoepidemiol Drug Saf*. (2006) 15:95–98. doi: 10.1002/pds.1171
29. Zong N, Kim H, Ngo V, Harismendy O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics*. (2017) 33:2337–44. doi: 10.1093/bioinformatics/btx160
30. Emig D, Ivliev A, Pustovalova O, Lancashire L, Bureeva S, Nikolsky Y, et al. Drug target prediction and repositioning using an integrated network-based approach. *PLoS ONE*. (2013) 8:e60618. doi: 10.1371/journal.pone.0060618
31. Ezzat A, Wu M, Li X, Kwoh CK. Computational prediction of drug-target interactions *via* ensemble learning. *Methods Mol Biol*. (1903) *pp*. (2019) 239–254. doi: 10.1007/978-1-4939-8955-3_14
32. He Z, Zhang J, Shi XH, Hu LL, Kong X, Cai YD, et al. Predicting drug-target interaction networks based on functional groups and biological features. *PLoS ONE*. (2010) 5:e9603. doi: 10.1371/journal.pone.0009603
33. Wang L, You Z-H, Chen X, Yan X, Liu G, Zhang W. RFDT: a rotation forest-based predictor for predicting drug-target interactions using drug structure and protein sequence information. *Curr Protein Pept Sci*. (2018) 19:445–54. doi: 10.2174/1389203718666161114111656
34. Huang Y, You Z, Chen X. A systematic prediction of drug-target interactions using molecular fingerprints and protein sequences. *Curr Protein Pept Sci*. (2018) 19:468–78. doi: 10.2174/1389203718666161122103057
35. Meng FR, You ZH, Chen X, Zhou Y, An JY. Prediction of drug-target interaction networks from the integration of protein sequences and drug chemical structures. *Molecules*. (2017) 22:1119. doi: 10.3390/molecules22071119
36. Xiao X, Min JL, Wang P, Chou KC. iGPCR-drug: a web server for predicting interaction between GPCRs and drugs in cellular networking. *PLoS ONE*. (2013) 8:e72234. doi: 10.1371/journal.pone.0072234
37. Tian K, Shao M, Wang Y, Guan J, Zhou S. Boosting compound-protein interaction prediction by deep learning. *Methods*. (2016) 110:64–72. doi: 10.1016/j.ymeth.2016.06.024
38. Hu PW, Chan KCC, You ZH. Large-scale prediction of drug-target interactions from deep representations. *Proc Int Jt Conf Neural Networks*. (2016) 1236–43. doi: 10.1109/IJCNN.2016.7727339
39. Wang L, You ZH, Chen X, Xia SX, Liu F, Yan X, et al. A computational-based method for predicting drug–target interactions by using stacked autoencoder deep neural network. *J Comput Biol*. (2018) 25:361–73. doi: 10.1089/cmb.2017.0135
40. Jin G, Wong STC. Toward better drug repositioning: prioritizing and integrating existing methods into efficient pipelines. *Drug Discov Today*. (2014) 19:637–44. doi: 10.1016/j.drudis.2013.11.005
41. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. (2017) 46:D1074–82. doi: 10.1093/nar/gkx1037
42. Fu G, Batchelor C, Dumontier M, Hastings J, Willighagen E, Bolton E. PubChemRDF: towards the semantic annotation of PubChem compound and substance databases. *J Cheminform*. (2015) 7:1–15. doi: 10.1186/s13321-015-0084-4

43. Bühlmann S, Reymond JL. ChEMBL-likeness score and database GDBChEMBL. *Front Chem*. (2020) 8:46. doi: 10.3389/fchem.2020.00046
44. Griffiths R-R, Jos' J, Miguel J, Hern'hernández-Lobato H. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chem Sci*. (2020) 11:577–586. doi: 10.1039/C9SC 04026A
45. Bjerrum EJ. SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules (2017). Available online at: https://arxiv.org/abs/1703.07076v2 (accessed January 13, 2022).
46. Scheife RT, Protein binding: what does it mean? *DICP*. (1989) 23:27–31. doi: 10.1177/106002808902300706
47. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. (2020) 49:D480–9. doi: 10.1093/nar/gkaa1100
48. Elabd H, Bromberg Y, Hoarfrost A, Lenz T, Franke A, Wendorff M. Amino acid encoding for deep learning applications. *BMC Bioinformatics*. (2020) 21:1–14. doi: 10.1186/s12859-020-03546-x
49. Ye H, Liu Q, Wei J. Construction of drug network based on side effects and its application for drug repositioning. *PLoS ONE*. (2014) 9:e87864. doi: 10.1371/journal.pone.0087864
50. Sonehara K, Okada Y. Genomics-driven drug discovery based on disease-susceptibility genes. *Inflamm Regen*. (2021) 41:1–5. doi: 10.1186/s41232-021-00158-7
51. Piñero J, Bravo L, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. (2016) 45:D833–9. doi: 10.1093/nar/gkw943
52. DisGeNET-a database of gene-disease associations. Available online at: https://www.disgenet.org/ (accessed January 13, 2022).
53. Piñero J, Saüch J, Sanz F, Furlong LI. The DisGeNET cytoscape app: exploring and visualizing disease genomics data. *Comput Struct Biotechnol J*. (2021) 19:2960–7. doi: 10.1016/j.csbj.2021.05.015
54. UniProt. Available online at: https://www.uniprot.org/ (accessed January 13, 2022).
55. Hubbard T. The Ensembl genome database project. *Nucleic Acids Res*. (2002) 30:38–41. doi: 10.1093/nar/30.1.38
56. Ensembl genome browser 105. Available online at: https://asia.ensembl.org/index.html (accessed January 13, 2022).
57. Kotz J. Phenotypic screening, take two. *Sci Exch*. (2012) 5:380. doi: 10.1038/scibx.2012.380
58. Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat Rev Drug Discov*. (2017) 168:16:531–43. doi: 10.1038/nrd.2017.111
59. Reaume AG, Drug repurposing through nonhypothesis driven phenotypic screening. *Drug Discov Today Ther Strateg*. (2011) 8:85–8. doi: 10.1016/j.ddstr.2011.09.007
60. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat*. (2012) 33:777–80. doi: 10.1002/humu.22080
61. Zheng W, Thorne N, McKew JC. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov Today*. (2013) 18:1067. doi: 10.1016/j.drudis.2013.07.001
62. Mungall CJ, McMurry JA, Köhler S, Balhoff JP, Borromeo C, Brush M, et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res*. (2016) 45:D712–22. doi: 10.1101/055756
63. El-Behery H, Attia AF, El-Feshawy N, Torkey H. Efficient machine learning model for predicting drug-target interactions with case study for COVID-19. *Comput Biol Chem*. (2021) 93:107536. doi: 10.1016/j.compbiolchem.2021.107536
64. Rifaioglu AS, Nalbat E, Atalay V, Martin MJ, Cetin-Atalay R, Dogan T. DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem Sci*. (2020) 11:2531–57. doi: 10.1039/C9SC 03414E
65. Validation of Predicting Drug-Food Interaction by BDDCS via a Literature Review. Available online at: https://www.researchgate.net/publication/308372028_Validation_of_Predicting_Drug-Food_Interaction_by_BDDCS_via_a_Literature_Review (accessed January 18, 2022).
66. PyMOL | pymol.org. Available online at: https://pymol.org/2/ (accessed January 18, 2022).
67. Seeliger D, De Groot LB. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J Comput Aided Mol Des*. (2010) 24:417–22. doi: 10.1007/s10822-010-9352-6
68. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. (2020) 49:D1388–95. doi: 10.1093/nar/gkaa971
69. Maier HJ, Bickerton E, Britton P. Coronaviruses: methods and protocols. *Coronaviruses Methods Protoc*. (2015) 1−282. doi: 10.1007/978-1-4939-2438-7
70. Shi H, Liu S, Chen J, Li X, Ma Q, Yu B. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*. (2019) 111:1839–52. doi: 10.1016/j.ygeno.2018.12.007
71. Rajput A, Thakur A, Mukhopadhyay A, Kamboj S, Rastogi A, Gautam S, et al. Prediction of repurposed drugs for Coronaviruses using artificial intelligence and machine learning. *Comput Struct Biotechnol J*. (2021) 19:3133–48. doi: 10.1016/j.csbj.2021.05.037
72. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3
73. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput Struct Biotechnol J*. (2020) 18:784–90. doi: 10.1016/j.csbj.2020.03.025
74. Nguyen DD, Gao K, Chen J, Wang R, Wei G-W. Potentially highly potent drugs for 2019-nCoV. *bioRxiv*. (2020). doi: 10.1101/2020.02.05.936013
75. FDA Approves First Treatment for COVID-19 | FDA. Available online at: https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19 (accessed January 25, 2022).
76. Tropmann-Frick M, Schreier T. Towards drug repurposing for COVID-19 treatment using literature-based discovery. (2022) 343:215–32. doi: 10.3233/FAIA210488
77. Monteleone S, Kellici TF, Southey M, Bodkin MJ, Heifetz A. Fighting COVID-19 with artificial intelligence. *Methods Mol Biol*. (2022) 2390:103–12. doi: 10.1007/978-1-0716-1787-8_3
78. Salgotra R, Rahimi I, Gandomi AH. Artificial intelligence for fighting the COVID-19 pandemic. *Humanit Driven AI*. (2022) 165–77. doi: 10.1007/978-3-030-72188-6_8
79. Hussain M, Jabeen N, Amanullah A, Ashraf Baig A, Aziz B, Shabbir S, et al. Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions. *AIMS Microbiology*. (2020) 6:350–60. doi: 10.3934/microbiol.2020021
80. Idris MO, Yekeen AA, Alakanse OS, Durojaye OA. Computer-aided screening for potential TMPRSS2 inhibitors: a combination of pharmacophore modeling, molecular docking and molecular dynamics simulation approaches. *Biomol Struct Dyn*. (2020) 39:1–19. doi: 10.1080/07391102.2020.1792346
81. Vankadari N, Structure of Furin Protease Binding to SARS-CoV-2 Spike Glycoprotein and Implications for Potential Targets and Virulence. *J Phys Chem Lett*. (2020) 11:6655–63. doi: 10.1021/acs.jpclett.0c01698
82. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Critical Care*. (2020) 24:422. doi: 10.1186/s13054-020-03120-0
83. Since January (2020). Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company' s public news and information. no. January (2020).
84. Zhao MM, Yang WL, Yang FY, Zhang L, Huang WJ, Hou W, et al. Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *Signal Transduction and Targeted Therapy*. (2021) 6:1–12. doi: 10.1038/s41392-021-00558-8
85. Sencanski M, Perovic V, Pajovic SB, Adzic M, Paessler S, Glisic S. Drug repurposing for candidate SARS-CoV-2 main protease inhibitors by a novel in silico method. *Molecules*. (2020) 25:3830. doi: 10.3390/molecules25173830
86. Weglarz-Tomczak E, Tomczak JM, Talma M, Burda-Grabowska M, Giurg M, Brul S. Identification of ebselen and its analogues as potent covalent

inhibitors of papain-like protease from SARS-CoV-2. *Sci Reports*. (2021) 11:1–10. doi: 10.1038/s41598-021-83229-6
87. Haritha CV, Sharun K, Jose B. Ebselen, a new candidate therapeutic against SARS-CoV-2. *Int J Surg*. (2020) 84:53. doi: 10.1016/j.ijsu.2020.10.018
88. Hijikata A, Shionyu-Mitsuyama C, Nakae S, Shionyu M, Ota M, Kanaya S, et al. Knowledge-based structural models of SARS-CoV-2 proteins and their complex with potential drugs. *FEBS Lett*. (2020) 594:1960–73. doi: 10.1002/1873-3468.13806
89. Kabir ER, Mustafa N, Nausheen N, Sharif Siam MK, Syed EU. Exploring existing drugs: proposing potential compounds in the treatment of COVID-19. *Heliyon*. (2021) 7:e06284. doi: 10.1016/j.heliyon.2021.e06284
90. Hydralazine Completed Phase N/A Trials for COVID / Coronavirus Disease (2019) (COVID-19) / High Blood Pressure (Hypertension) | DrugBank Online. Available online at: https://go.drugbank.com/drugs/DB01275/clinical_trials?conditions=DBCOND0129755%2CDBCOND0127821%2CDBCOND0040789&status=completed (accessed January 25, 2022).
91. Yalçin S, Yalçinkaya S, Ercan F. Determination of potential drug candidate molecules of the hypericum perforatum for COVID-19 Treatment. *Curr Pharmacol Reports*. (2021) 7:42–8. doi: 10.1007/s40495-021-00254-9
92. Lu J, Hou Y, Ge S, Wang X, Wang J, Hu T, et al. Screened antipsychotic drugs inhibit SARS-CoV-2 binding with ACE2 *in vitro*. *Life Sci*. (2021) 266:118889. doi: 10.1016/j.lfs.2020.118889
93. Uckun FM, Saund S, Windlass H, Trieu V. Repurposing anti-malaria phytomedicine artemisinin as a COVID-19 drug. *Front Pharmacol*. (2021) 12:407. doi: 10.3389/fphar.2021.649532
94. Cheng F, Murray JL, Zhao J, Sheng J, Zhao Z, Rubin DH. Systems biology-based investigation of cellular antiviral drug targets identified by gene-trap insertional mutagenesis. *PLoS Comput Biol*. (2016) 12:1–27. doi: 10.1371/journal.pcbi.1005074
95. Alkotaji M, Azithromycin and ambroxol as potential pharmacotherapy for SARS-CoV-2. *Int J Antimicrob Agents*. (2020) 56:106192. doi: 10.1016/j.ijantimicag.2020.106192
96. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, BarabáSi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun*. (2018). 9:1–12. doi: 10.1038/s41467-018-05116-5
97. Noori M, Nejadghaderi SA, Sullman MJM, Carson-Chahhoud K, Kolahi AA, Safiri S. Epidemiology, prognosis and management of potassium disorders in COVID-19. *Rev Med Virol*. (2022) 32:e2262. doi: 10.1002/rmv.2262
98. Ezzat A, Wu M, Li XL, Kwoh CK. Drug-target interaction prediction via class imbalance-aware ensemble learning. *BMC Bioinformatics*. (2016) 17:267–76. doi: 10.1186/s12859-016-1377-y
99. Gharagozloo M, Suchan M, Kaliarik L, Krempaska S, Koval J. CLINICAL STUDY Immune-mediated cochleovestibular disease. *Bratisl Lek Listy*. (2015) 116:296–301. doi: 10.4149/BLL_2016_017
100. Ezzat A, Wu M, Li XL, Kwoh CK. Drug-target interaction prediction using ensemble learning and dimensionality reduction. *Methods*. (2017) 129:81–8. doi: 10.1016/j.ymeth.2017.05.016
101. Abdel-Bakky MS, Amin E, Faris TM, Abdellatif AAH. Mental depression: relation to different disease status, newer treatments and its association with COVID-19 pandemic (Review). *Mol Med Rep*. (2021) 24:1–17. doi: 10.3892/mmr.2021.12479

Conflict of Interest: JL and YK were employed by BioSpero. KC is owner of BioSpero.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ahmed, Lee, Samantasinghar, Kim, Kim, Kang, Memon, Lim and Choi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fpubh.2022.893989

Perception Exploration on Robustness Syndromes With Pre-processing Entities Using Machine Learning Algorithm

Pravin R. Kshirsagar[1], Hariprasath Manoharan[2], Shitharth Selvarajan[3], Hassan A. Alterazi[4], Dilbag Singh[5] and Heung-No Lee[5]**

[1] Department of Artificial Intelligence, G.H. Raisoni College of Engineering, Nagpur, India, [2] Department of Electronics and Communication Engineering, Panimalar Institute of Technology, Chennai, India, [3] Department of Computer Science and Engineering, Kebri Dehar University, Kebri Dehar, Ethiopia, [4] Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, [5] School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Mohamed Sirajudeen, VIT-AP University, India
Muralidharan C, SRM Institute of Science and Technology, India

***Correspondence:**
Heung-No Lee
heungno@gist.ac.kr
Hariprasath Manoharan
hari13prasath@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *11 March 2022*
Accepted: *27 April 2022*
Published: *16 June 2022*

Citation:
Kshirsagar PR, Manoharan H, Selvarajan S, Alterazi HA, Singh D and Lee H-N (2022) Perception Exploration on Robustness Syndromes With Pre-processing Entities Using Machine Learning Algorithm. Front. Public Health 10:893989. doi: 10.3389/fpubh.2022.893989

The majority of the current-generation individuals all around the world are dealing with a variety of health-related issues. The most common cause of health problems has been found as depression, which is caused by intellectual difficulties. However, most people are unable to recognize such occurrences in them, and no procedures for discriminating them from normal people have been created so far. Even some advanced technologies do not support distinct classes of individuals as language writing skills vary greatly across numerous places, making the central operations cumbersome. As a result, the primary goal of the proposed research is to create a unique model that can detect a variety of diseases in humans, thereby averting a high level of depression. A machine learning method known as the Convolutional Neural Network (CNN) model has been included into this evolutionary process for extracting numerous features in three distinct units. The CNN also detects early-stage problems since it accepts input in the form of writing and sketching, both of which are turned to images. Furthermore, with this sort of image emotion analysis, ordinary reactions may be easily differentiated, resulting in more accurate prediction results. The characteristics such as reference line, tilt, length, edge, constraint, alignment, separation, and sectors are analyzed to test the usefulness of CNN for recognizing abnormalities, and the extracted features provide an enhanced value of around 74%higher than the conventional models.

Keywords: convolutional neural network (CNN), depression characteristics, emotion recognition, machine learning (ML), perception

INTRODUCTION

Teenage mental illness is a major psychological condition that causes constant emotional distress as well as a loss of enthusiasm and energy in accomplishing any work. It has an impact on teenage thoughts, feelings, and behavior characteristics, resulting in emotional failure with functional and personal concerns. Teenage pupils who are experiencing depression have significant problems carrying out daily tasks and performing in various situations.

According to a poll (1) more than 21% of teenagers in India reported suffering from serious depression in the previous year. Friendships and emotional bonds between families and friends are harmed by depression. It has a potential to slow down or halt the growth of the children. As a result of this, suicidal thoughts and attempts may result. Handwriting analysis, video and audio analysis, gesture and movement analysis, and psychological evaluation are among the many different approaches for detecting depression and other mental illnesses that might be created utilizing machine learning (ML) and deep learning methodologies (1). Depression is also one of the primary causes of the disability in teens. Often, a person who is going through this is hesitant to seek help from a counselor. In this case, technology can play a critical role and be the first line of defense. As people become more active on social media, their patterns can be exploited to identify physical and mental illnesses in simpler manner. Several ML algorithms for regression, such as random forest, linear regression, Bayesian network, Naive Bayes, support vector machine (SVM), *k*-nearest neighbor, and deep learning algorithms for feature extraction, such as Convolutional Neural Networks (CNN), have been used and successfully tested on various datasets containing video, audio, texts, and images fed to a classifier, applicable on affective and personality computing with feature analysis.

In the previous eras, there was an increase in the use of media information in an online method for recognizing, measuring, and tracking changes in the event of an illness. The universal design of web-based media presents an opulent case for improving the data available to psychological wellness therapists and analysts, allowing for an education expert and a well-organized expressive comfort field. Furthermore, entities are influenced by communicable cynical attitudes in informal associations, which stimulate dejection and other psychological diseases. Dysfunctional conduct is known to be a substantial risk factor for self-destruction; about 80% of those who attempt or die by self-destruction have been diagnosed with some form of psychological instability. Wretchedness is the most well-known psychological disorder, although it has remained unknown or untreated due to its recognition or denial. Early recognition of a significant sorrow's signs and therapy through optimal mediation can help to prevent the onset of a significant sorrow. Numerous studies have identified physical and psychological disorders as a result of the vast amounts of data available on the internet, including a few studies that focused specifically to depression.

LITERATURE SURVEY

Anxiety, despair, stress, and other mental health concerns occur too frequent in youth. It is critical to detect such mental diseases earlier to avoid future effects. Currently, numerous methodologies based on AI-based approaches and algorithms for detecting depression and other mental diseases have been developed. The researchers created a ML model that analyzes social media posts and tweets to detect depression. The researchers used data gathered from social media profiles, activity, and tweets to train their model for this project. The test dataset findings reveal that the critical values for classification error metrics are determined by the sample-based features collected from social media analysis. The model's accuracy will increase as the number of features increases. This is a data-driven paradigm in which the performance and outcomes are determined by data and characteristics (2). The researchers aimed to develop a method for examining facial features and detecting depression scales using image processing techniques. They used characteristics and facial expressions to train the classifier for pleasant and bad emotions. A teen student's video is collected and sent into the system. Using Gabor filters, the model extracts the face and its features. The collected characteristics are then classified into different classes using an SVM technique with supervised learning classifier classification. The number and severity of depressive patients are computed using a neural network to analyze depressive emotions taken from a video sample. The class coordinator and parent are advised about the student's disturbed mental situation after the depression is calculated using the Beck's Depression Inventory (BDI-II) scale. Other characteristics used by the system include academic performance, extracurricular performance, and conduct opinion (3).

Deep learning algorithms have a wide range of applications in the treatment and analysis of mental health and other mental disorders to better identify biomarkers relevant to unique characteristics between them, and they can be better identified in a variety of situations, such as mental depression, where no models exist. Furthermore, the state of sadness is described as an event sequence, allowing users to compute psychological infection or severe ML circumstances. Traditional researches have stated that generated models are implemented with strategic tools that pre-process MRI data, such as MRI scanning tomography, for extracting all independent functionalities from deep learning models that are used as input data to validate the model. All of these aforementioned gadgets, on the other hand, collect data based on the prior knowledge of the reference set. As a result, certain datasets that might be used to predict mental disease or healthy cases of depression may be misplaced. Cooperation was incorporated in the analysis to improve the accuracy of the outcomes of this multi-modal investigation. Researchers worked with the National Institute of Neurological Disorders and Stroke, the Centers for Disease Control and Prevention (CDC), and the American Psychiatric Association (APA) to aggregate the data, in addition to the CAN–BIND study. This combined effort could hopefully result in easily available biological assays that can accurately and efficiently guide treatment selection for depression (4). Daily chores such as drawings or writings might be used to detect depression, anxiety, or any other mental problems. Emotional State Recognition from Handwriting and Drawing (EMOTHAW) is a public database that contains handwriting and drawing samples pertaining to an individual's emotions and mental state. This collection contains handwritten and drawn samples from 129 people who are facing mental and emotional problems. The participants were given the task of sketching a pentagon, a house, a circle, and a clock, as well as writing a cursive statement. The elements from the sample are used to analyze a person's personality and emotional

aspects, resulting in a personality and character profile. Starting point, gradient, size, border, weight, spacing, strokes, and areas are some of the features analyzed from the handwriting sample that distinguishes it from other samples. The results of the sample analysis were uploaded to their database. To classify and evaluate the samples, the researchers employed a random forest approach. This ML approach is based on an ensemble-based decision tree learning, which includes a feature rank procedure. To study an individual's mental state, feature ranking is used to rank and find the key elements in the sample (5–18).

Artificial intelligence (AI) (i.e., ML) has been introduced into the therapeutic field as a means of improving exactness and accuracy while reducing the number of time-consuming tasks that require human intervention (19). There is emerging evidence that ML's innovation might potentially discover and considerably enhance therapy of complicated mental problems such as despair (20). The researchers devised a system capable of detecting doom with minimal human intervention: man-made reasoning mental assessment. This structure consists of a brief human-PC intelligent assessment that employs artificial consciousness, namely, profound learning, to predict if a member would be deterred by palatable execution. Because of its ease of use, this innovation could provide a useful tool for mental health professionals to spot signs of distress, allowing for quicker intervention. Furthermore, by providing a more targeted assessment, it may minimize the difficulty of identifying and decoding profoundly complicated the physiological and social biomarkers of misery (6). Based on the patient's condition and the doctor's viewpoint, a comprehensive literature review is conducted. Visits to various clinics are used to obtain information. With Doctor's consent, information is accessed and reviewed (21, 22). While conducting a depression analysis, several fields of doctors' opinions on depression are examined. Types and causes of depression are investigated to develop a solution for prediction (17, 23–27).

Depression is a psychiatric condition that obstructs a person's growth by causing behavioral, emotional, and mental alterations.

TABLE 1 | Model Summary for feature extraction using Handwriting and Drawing sample.

Layer types	Output shape	Parametric values
2D Convolutional (Conv2D)-5	0, 222, 222, 32	898
2D Convolutional-6 (Conv2D)	0, 220, 220, 64	18,494
2D Maximum_pooling-4	0, 110, 110, 64	0
Dropout-5 (Dropout)	0, 110,110,64	0
2D Convolutional-7 (Conv2D)	0, 108,108,64	36,926
Dropout-6 (Dropout)	0,54,54, 64	0
2D Convolutional-8 (Conv2D)	0, 52,52,128	73,858
Dropout- 8	0,26, 26, 128	0
Flatten- 3	0, 86528	0
Dense- 4	0, 64	5,537,858
Dropout-7	0, 64	0
Dense-5	0, 1	65

Due to the lack of marking and measuring standards for the depression scale, the psychiatrist may be unable to recognize it. Electroencephalogram (EEG) signals can be used to identify detrimental alterations in the cerebrum's functioning. Thus, the suggested method can scale the downturn measure and recognize the degree of the illness using EEG data using AI methodologies. The highlights from the cerebrum's transitory region are isolated using six (FT7, FT8, T7, T8, TP7, and TP8) channels in this study. Delta, theta, alpha, beta, gamma1, and gamma2 band power, as well as their associated unevenness and matching lopsidedness, are the straight highlights used. Sample entropy (sample) and detrended fluctuation analysis (DFA) are two non-straight highlights used. The following classifiers were used: SVM bagging alongside three distinct piece capacities (polynomial, Gaussian, and sigmoidal). Relief-F is the highlight selection approach that was used. Using SVM (Gaussian Kernel function) and Relief-F as a highlight option, the most notable arrangement precision of 96.02 and 79.19% was achieved for the location and seriousness scaling of darkness. The investigation revealed that a decline has an impact on the worldly portion of the mind (temporo–parietal region). It is indicated that the presence of downturns in the system is having the capability to directly affect the equator reference band structures. It may also be deduced that, among the many components of SVM, the Gaussian bit is more capable than other bits. The combination of all combined unevenness and deviation revealed exceptional arrangement precision (exactness of 90.26% for depression discovery and exactness of 75.31% for seriousness scaling) (7) among the many highlights. Normally, determining gloom necessitates a thorough examination by a knowledgeable expert. Recently, much more thoughts have been devoted to programmed melancholy expectation to conduct a more thorough and fruitful examinations of unhappiness. In this research, a unique approach is proposed that uses a two-stream profound spatial–temporal architecture to estimate the level of downturn from video data. The Inception-ResNet-v2 network is used in our method to extract the spatial data. A volume neighborhood directional number is described, which is based on a powerful component description to capture facial movements. Then, the element map obtained from the Very-low-density lipoprotein (VLDN) is applied to a convolutional neural network to obtain more discriminative highlights. Furthermore, by combining the worldwide middle pooling strategy into the model, a multi-layer bidirectional long transient memory model is planned to obtain transitory data. On the transient parts of spatial and worldly highlights, the TMP method is used. Finally, a thorough study of two testing datasets, AVEC-2013 and AVEC-2014, demonstrates that the proposed strategy outperforms the current methodologies for predicting suffering levels (8). **Table 1** indicates the Model Summary for feature extraction.

Research Gap and Motivation

The major objective of the proposed work is to overcome the disadvantages that exists in the prediction model using handwritten statements where multiple formats are not present is any existing models (1–29). In addition, only occasional treatments are provided for predicting depression with low complexity algorithms. Even the researchers have examined

the implementation model with handwritten statements where complexity of prediction is much higher. Moreover, in the presence of electrograph signals, it is much easier to predict different characteristics of depression but an electrical utensil is needed in this kind of detection. This in turn will create several back-end problems in the near future and the data gathering approach is much reduced in processing of signals. Therefore, to overcome the abovementioned gap, the proposed method on depression prediction using CNN has been implemented.

Proposed Methodology

The features derived from the output samples are used to analyze a person's personality and emotional components, resulting in a personality and character profile. Handwriting samples are examined for the properties such as reference line, tilt, length, edge, constraint, alignment, separation, and sectors that distinguish them from other samples. Handwriting samples of teens are collected and processed using a computerized deep learning utilizing the BDI-II scale. Samples are classified according to their rigorousness in the suggested model architecture to identify, scale, and classify depression using handwriting samples of teenagers. This study examines several depression scores and categories, such as anxiety or stress from 0 to 13, mild depression/anxiety from 14 to 19, moderate depression/anxiety from 20 to 28, and severe depression/anxiety from 29 to 63. In addition, for analyzing handwriting photographs, CNN is used with image recognition techniques to extract deep features from the sample. The CNN, or advanced neural network layers, is a type of deep learning technology that has been effectively used to solve real-world problems.

The network architecture is set up in such a way that extracting and analyzing features from images, audio signals, text, and video segments can be done quickly and efficiently. The proposed method on depression prediction is implemented using CNN for extracting all necessary features by processing the captured input images. In addition to input images, a handwritten sample is trained using hidden layer approach. This type of processing technique provides additional advantage as all dataset features are obtained using compatible mode of operation even if the images are provided at low resolutions. However, the images are converted to high resolution in the CNN model thus data is gathered after classification mechanism. All the steps are implemented under small time scale where errors are minimized at corresponding stages. Furthermore, during feature extraction, the accuracy of detection using CNN is maximized using rotation and scaling techniques.

Objectives

The proposed method on predicting different characteristics of depression aims to solve a multi-objective case study that can be described as follows:

- To minimize the errors that are present in existing models using CNN and maximize the accuracy of prediction by implementing two different sample sets.
- To implement BDI-II scaling for images during classification and extraction techniques where high-resolution images can be processed.
- To integrate cloud storage for storing all monitored parametric values using reference line measurements.

Study Structure

The remainder of this study is laid out as follows. The proposed work's design technique is discussed in Section Design Methodology. The feature categorization technique is discussed in Section Optimization Algorithm: Classification and Regression Development with a step-by-step implementation. The results and overall consequences are discussed in Section Results and Discussions, and the importance of the projected model is described in Section Conclusions.

DESIGN METHODOLOGY

A model has been constructed in this architecture to detect an individual's emotional state using handwriting and drawing examples. The model starts with handwriting and drawing samples as input, which it runs through a CNN-based classifier. A database is created for this system by collecting handwriting text and drawing samples from 250 people. A separate work was also provided, which required students to draw a pentagon, a house, a circle, and a clock, as well as write a statement in cursive. Individuals were given a tablet on which to draw and write for the assignment indicated in **Figure 1**. The samples are stored in a database once the process is completed. In addition, the database is separated into the following three categories: Training, testing, and validation, with an 80:10:10 ratio.

Various characteristics of drawing and handwriting samples, such as reference line, tilt, length, edge, constraint, alignment, separation, and sectors, that distinguish one sample from another are examined. Wide separation between words/lines, increasing benchmark, unbalanced left and right edges, and less top edge were a few of the overall highlights of the database. The most commonly observed features for scale are as follows: Maximum weight, broadness at left-edge and top-edge, blended inclination, broadness dispersing in the middle of the lines, maximum slanting in letters such as F, G, H, K, and R, maximum slanting in letters such as F, G, H, K, and R, maximum slanting in letters such as F, G, H, K, and R, maximum slanting in letters such as F, G, H, K, and R. The features are accurate inclination, wide separation between lines/characters, and blended gauge on a scale ranging from 14 to 19, which are collectively defined as a mild depression/anxiety class. The highlights of the lopsided gauge, lopsided edge, long bending within alphabets such as Y, G, F, and so on, enormous dimension of alphabets, wideness in the left-edge, and broad dispersing in the middle of sentences for scale 20–28 are defined as a moderate depression/anxiety class. The top-most and leftmost edge wideness, blended/leftmost inclination, falling gauge, and asymmetrical dispersal in the midst of the sentences are the highlights for the nervousness–anxiety class. Maximum weight, rightmost inclination, blended standard,

small size, rising benchmark, lopsided edge, huge bends with alphabets such as S, G, Y, F, and so on. are seen in the sadness–nervousness–stress class, while odd edges, linear gauge, linear/rightmost/leftmost inclination, and average sizing are seen in typical classes.

Handwriting samples were collected and used as a tool for validation, assessing intellectual deficiency, and assessing character attributes. In truth, collecting handwriting samples is non-intrusive, simple, and low-tech, and requires no administrative expertise. The following samples are saved in

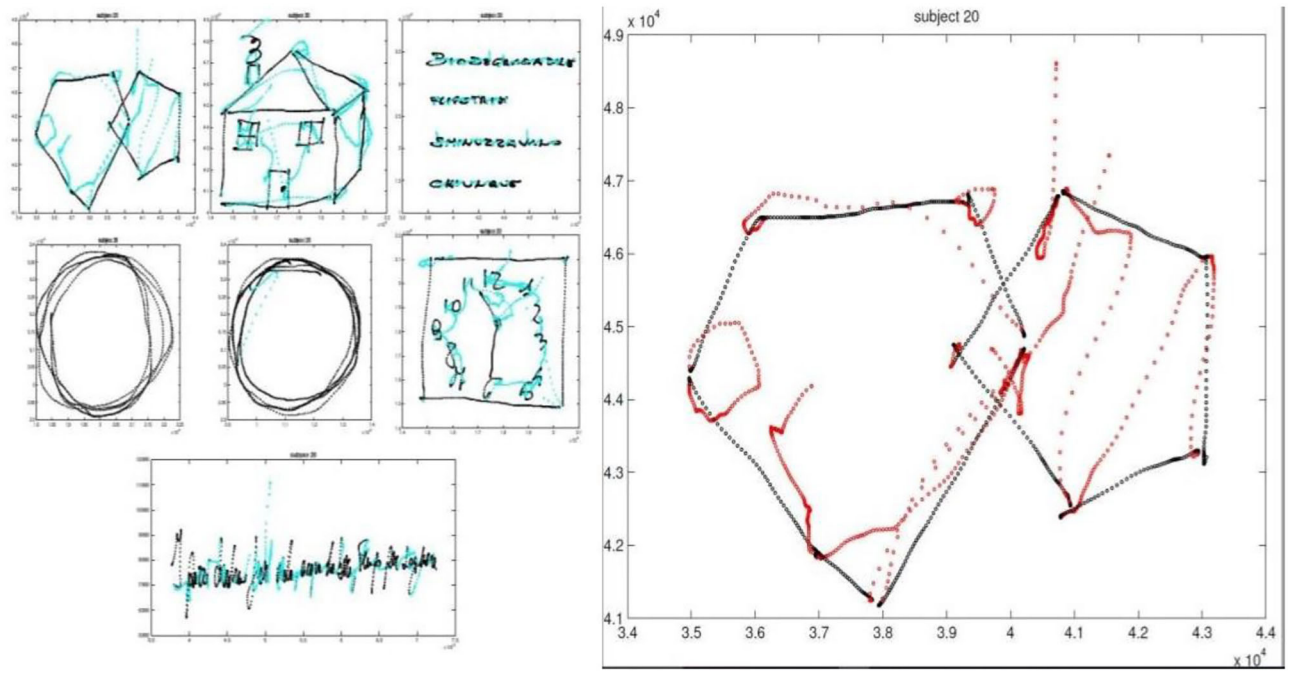


FIGURE 1 | A sample task to draw a pentagon, a house, a circle and a clock and text in cursive.

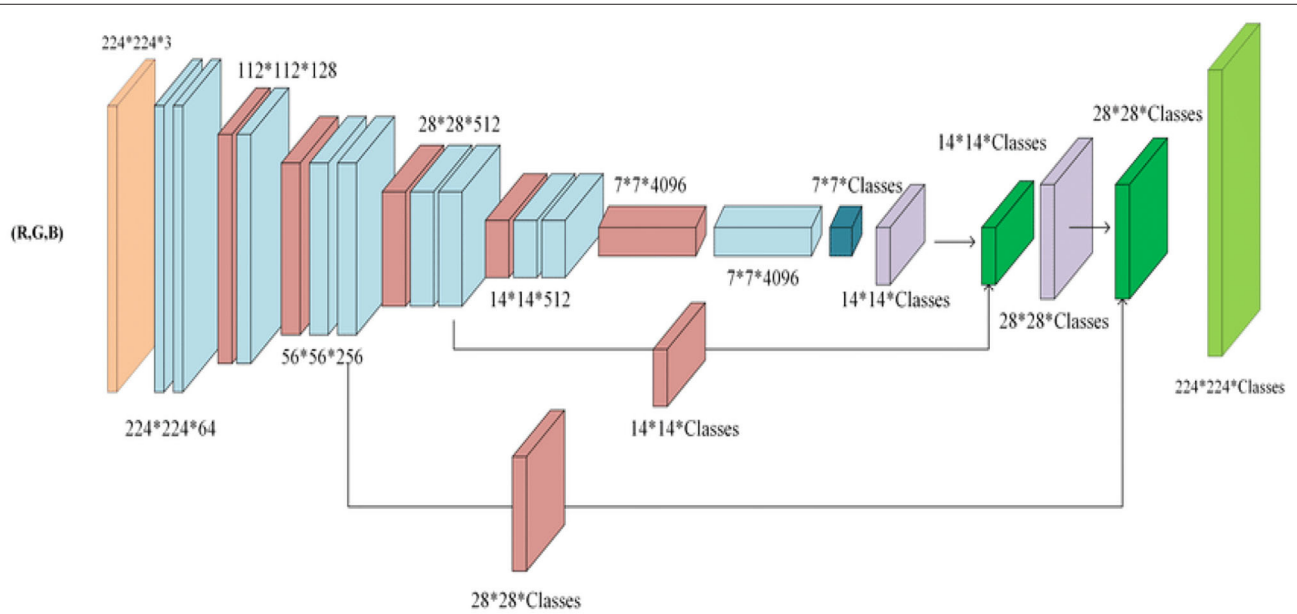


FIGURE 2 | Network architecture for eight-layered CNN.

Switched Virtual Circuit (SVC) records, where SVC stands for record expansion. The SVC extension documents are typically presented in standard American Standard Code for Information Interchange (ASCII) formats that can be accessed using applications such as Microsoft Word, Google Docs, and other common trending applications. The following benchmarking parameters are applied to the data samples:

- Location in x-hub.
- Location in y-hub.
- Time stamp.
- Pen up = 0 or Pen down = 1.
- Dimension of the pen as for the tablet.
- Dimension of the pen as for the tablet.
- Pen point pressure applied by the pen.

Additional data, such as speed highlights (quickening, speed), prompt direction point, brief dislodging, time highlights, and features highlights, can be produced using this arrangement of dynamic data.

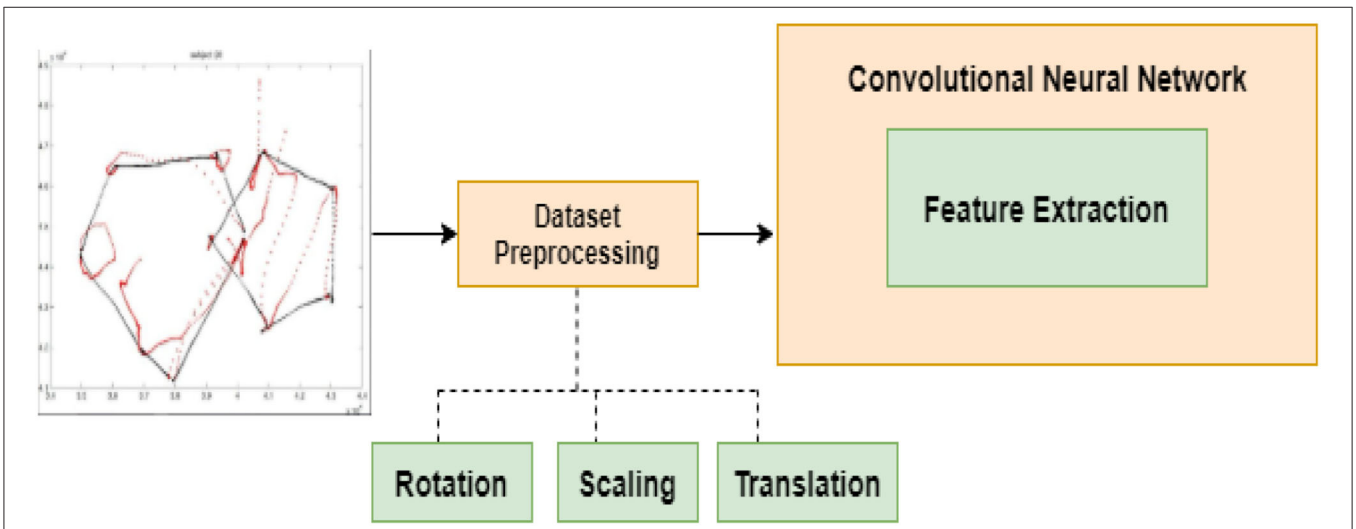


FIGURE 3 | Feature extraction for drawing and handwriting samples.

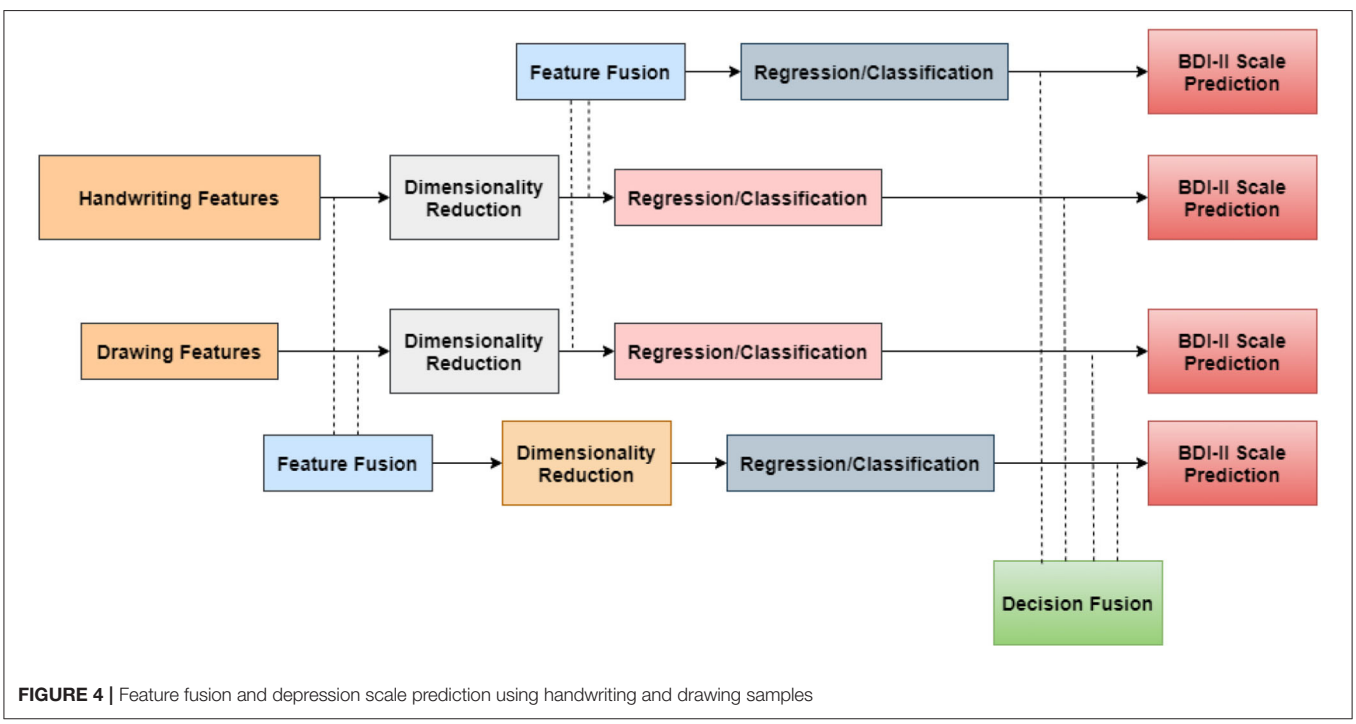


FIGURE 4 | Feature fusion and depression scale prediction using handwriting and drawing samples

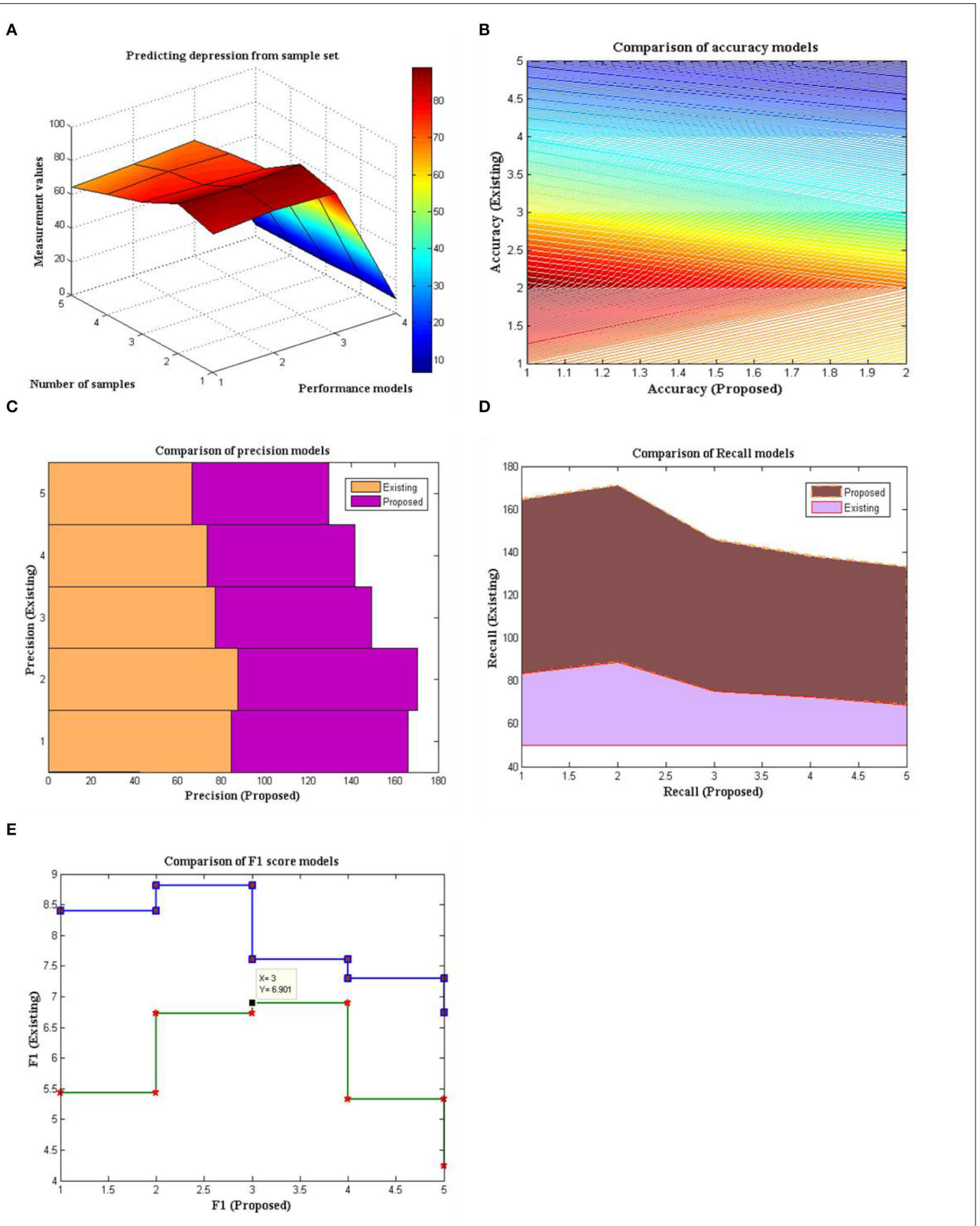


FIGURE 5 | Sample set vs. Measurement values **(A)** Total comparison, **(B)** accuracy, **(C)** precision, **(D)** recall, and **(E)** F1-Score.

Feature Extraction From Handwriting and Drawing Samples

The generated dataset is pre-processed by rotating, scaling, and translating it in order. The photos are resized and normalized to be compatible with the system. This dataset was used to train the convolutional neural network model. Following training, a new input image is given to the neural network to extract features and predict the BDI-II scale, as shown in **Figure 2**.

The component maps are used to lower the dimensionality of the convolutional layers using the created pooling layer. The framework executes the maximum pooling layer into the properly defined element map that are unpredicted upon the indicated size, and the largest component is chosen from the framework. There is a complete layer bound after the mixing of various layers, which bundles the highlights to address questionable classes utilizing the increased level features from the Convolutional Layer/Pooling Layer. Soft-max functions are engaged as the last layer, which fixes multiple yields from each unit to the features from beginning to final values. By assessing the vulnerability at each layer, the total cross-entropy job analyses the presentation of collecting ideals. The error that is corrected at the yield layer must be reduced.

The reverse engendering method, which is carried out by the inclination drive technique, attempts to reduce the error by progressing regressively over distinct levels by modifying multiple assessments of loads and tendencies. A cross-approval approach has been used in the development model for evaluating AI models by preparing a group of datasets and measuring using correlative separations on the dataset. The *k*-cover cross-endorsement condenses data into *k* bits of equal dimensions, with one segment adding separately for endorsement and the rest sections used for readiness in this model. The entire dataset is split into *k* wrinkles of similar size, and the model is built using $k - 1$ folds. The parametric test of the model on the *k*th overlay will begin immediately and continue in the foreseeable future. For each cover, the goof/setback, exactness is set aside. This cycle is repeated until all of the creases in the testing dataset have been applied. Accuracy stored in *k*-folds is used to resolve the typical exactness. By combining a few models, cluster knowledge representation facilitates the application of different representations in AI. A basic research control is executed a couple of times in a bunching calculation to outline for choosing from the hypotheses that are gathered. Cross-approval was used by placing persons with similar ensemble features on different planning circles and using different models for each participant in the approval process. Outfits are commonly used to obtain claimed categories that are obtainable from a collection of arrangements that produce low-precision results. As a result, the multiple layers that are used as a fundamental classifier in the proposed work will assemble the exactness after cross-approval an outfit methodology is functional where the top classifier loads out of the eight convolutional neural organization models that are used while making a gathering model as depicted in **Figure 3**.

OPTIMIZATION ALGORITHM: CLASSIFICATION AND REGRESSION DEVELOPMENT

The process of predicting depression from textual data has been implemented with high-effect analysis using long short-term memory which is indicated as productive model (30). In this model, a handwritten content is placed and by using the frequency the deep learning model predicts the true and false positivity rates. However, using the frequency of recognition accuracy of content cannot be achieved to large extent. Therefore, to achieve high accuracy many challenges with ML techniques are deliberated with suitable application platforms (31). One best application for ML technique is depression prediction where all possible problems that affects the mental health of humans are discussed (31). However, a comparative analysis will not provide solutions to overcome the problems in health-related issues which indicates that a testing model must be placed during comparison state. To have a deep insight knowledge testing model comparisons are made (32) using long short-term memory in two different directions. In this bidirectional procedure, the neural networks are incorporated and the visualization approaches are built for extracting high-feature comparisons. After extracting the characteristics from the handwriting, drawing, and audio samples, the depression scale is predicted using several ML methods, and a comparison is made based on the results. The findings of this comparative research will provide the best performing algorithm for obtaining correct results. Random forest, logistic regression, Naive Bayes, *k*-nearest neighbor, and SVM methods are used to calculate the BDI-II scale using regression techniques. Comparative analysis is used to further discuss the algorithms.

A range of elements have been derived from distinct as well as the same modality in this design, as seen in **Figure 4**. Deep neural networks are used to extract information from handwriting and artwork samples. In addition, a fusion method was used to integrate the various retrieved feature sets. After the features are extracted, a fusion approach is used to combine feature vectors from various modalities in a sequential order. This sequence was repeated twice—once before and once after the dimensionality reduction. The fusion at the decision-making level necessitates the clustering of prediction or classification findings. Based on the extracted features, the model's prediction results can be dependent on a single or the multiple modalities. The fusion in regression is considered to be difficult to use due to the existing problem with multi-collinearity, which is used to characterize the observational features from the dataset used for assessing depression class and scale. As a result, in the multi-collinearity model for binary classification, logical operations and logical gates are used to apply the decision fusion technique. Even logics are utilized to detect the presence or absence of depression and categorize them. In addition, the participants are categorized using logic operators AND and OR to fuse the classification findings. A random forest algorithm is utilized as a classifier in this methodology, which is a common way for decision fusing classifier algorithm to train parameterized scores signifying

TABLE 2 | Comparative analysis of different ML algorithms used for predicting depression from samples.

Models	Accuracy	Precision	Recall	F1-Score
SVM	82.54	84.54	83.36	8.394
Random forest	88.97	87.46	88.77	8.811
Naive Bayes	78.21	76.96	75.12	7.6028
k-Nearest neighbor	71.36	73.49	72.48	7.2981
Logistic regression	64.52	66.16	68.83	6.744

TABLE 3 | Error values of different ML algorithms used for predicting depression from samples.

Models	RMSE	Mean average error	Mean error
SVM	8.8548	8.8942	9.6363
Random forest	6,541	8.4235	8.5896
Naive Bayes	10.2254	9.2152	10.2203
k-Nearest neighbor	4,141	8.4458	10.9299
Logistic regression	12.6656	11.5569	12.6112

distance. The random forest classifier's output is utilized as an input for the next level regressors. In the case of observation utilizing regression over choice fusion, the generalized linear models are used, and expectation maximization procedures are applied to the decision fusion level, maximizing the possibilities. The major intention for choosing CNN is that in proposed method an automatic detection is needed where all features must be extracted within a short period of time. Also, to avoid human interference as system must be chosen that learns the descriptive model by its own and high computational effects can only be observed if CNN is implemented. In addition, the location of implementation is much complex in other algorithms but in CNN the implementation process can be done without any reference location thus making the model much easier during training stages. Further, the reliability of operation is much higher with a few label sets from the operating ends; thus, CNN is preferred for predicting the depression state using written sets.

Design Algorithm for Handwriting and Drawing Feature Extraction

***Step 1*:** Divide the dataset in 80:10:10 for training, testing, and validation.

***Step 2*:** Using feature ranking, set the weights and features as key parameters.

***Step 3*:** On the input image, do scaling, dimensionality reduction, rotation, and translation.

***Step 4*:** Carry out the pooling and forward propagation operations.

***Step 5*:** Using the Softmax activation layer function, calculate probabilities for each class.

***Step 6*:** Determine the model's output error.

***Step 7*:** The inclinations of the error are determined in the back propagation step by taking the incomplete subordinate for all loads. For limiting the yield blunder, the boundary values are refreshed.

***Step 8*:** For each image in the training dataset, repeat steps V through IX.

***Step 9*:** Using the provided equations, calculate the model accuracy and model loss for the proposed model for the validation dataset.

***Step 10*:** Save the optimal weights for future computations and modeling.

***Step 11*:** Create an ensemble-based learning model using cross-validation modeling.

***Step 12*:** Using the provided equations, calculate the model accuracy and model loss for the proposed model for the testing dataset.

Model Training and Evaluation of CNN

The accuracy for prediction is calculated using Eq. (1) as follows:

$$A_i = \left(\frac{CP_t(i)}{s_t(i)}\right) * 100 \tag{1}$$

where $CP_t(i)$ and $s_t(i)$ denote the total number of correct estimates and models, respectively.

The level of absolute number of right predictions to the all Outage tests is determined as accuracy. In the event that the model has a precision rate nearest to 1, it is considered as the model has more precise expectations and if the precision rate is nearest to 0, the model is viewed as less precise. The accuracy in terms of classification can be calculated as follows:

$$A_c\,(i) = \sum_{i=1}^{n} \left(\frac{tp\,(i) + tn(i)}{tp\,(i) + tn\,(i) + fp\,(i) + fn(i)}\right) \tag{2}$$

Here, $tp\,(i)$, $tn\,(i)$, $fp\,(i)$, and $fn\,(i)$ denote true positive, true negative, false positive, and false negative values, respectively.

Hence the values for sensitivity can be calculated using true positive and false negative values as follows:

$$\rho_i = \sum_{i=1}^{n} \frac{tp(i)}{tp\,(i) + fn(i)} \tag{3}$$

The value for specificity is also measured using true negative and false positive values as follows:

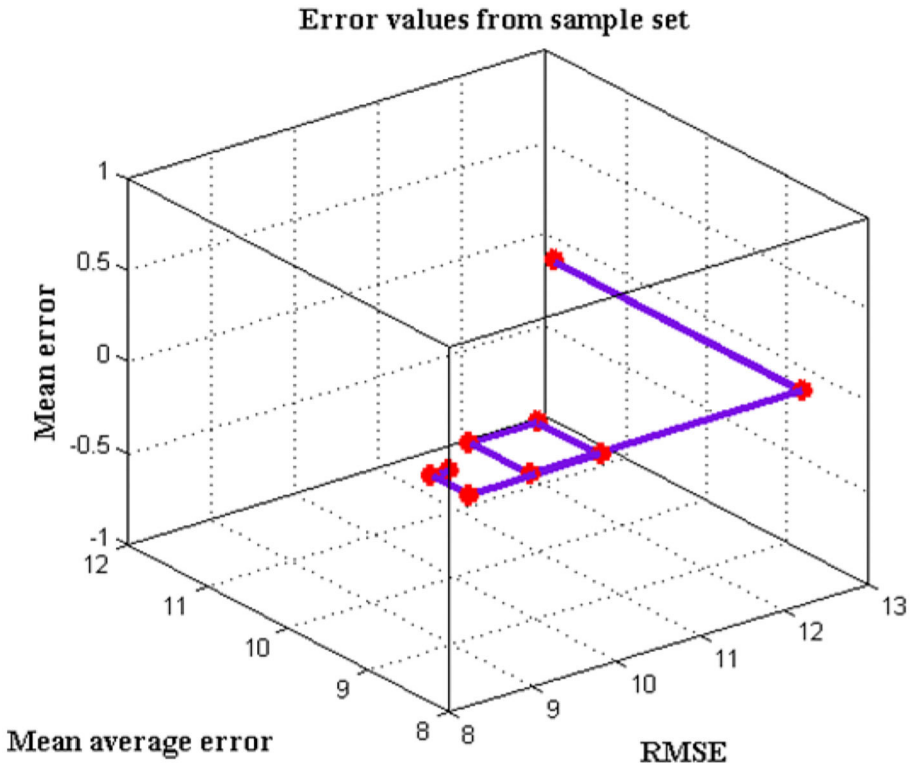


FIGURE 6 | Error values attained from sample dataset.

$$\sigma_i = \sum_{i=1}^{n} \frac{tn(i)}{fp\,(i) + tn(i)} \tag{4}$$

Similarly, the positive predictive value and negative predictive value for the proposed prediction model can be calculated using Eq. (5) as follows:

$$PV_p\,(i) = \sum_{i=1}^{n} \frac{tp(i)}{tp\,(i) + fp(i)} \tag{5}$$

$$PV_n\,(i) = \sum_{i=1}^{n} \frac{tn(i)}{tn\,(i) + fn(i)} \tag{6}$$

To evaluate the percentage error in predicting the root mean square error values are designed using the depressed and non-depressed classes by the model and thus calculating the error difference with the actually predicted values. The root mean square Error (RMSE) and average estimation error (AEE) can be calculated as follows:

$$RMSE_i = \sqrt{\frac{\sum_{i=1}^{n} (e_i - r_i)}{Total\ number\ of\ samples}} \tag{7}$$

where, e_i and r_i denote expected and reference values, respectively.

$$AAE_i = \sum_{i=1}^{n} \left| \hat{z_{in}} - z_i \right| \tag{8}$$

where $\hat{z_{in}}$ and z_i denote difference between class errors.

In analytical terms, the objective function can be defined using Eqs. (9) and (10), where the unconnected non-linear equations are established for both minimization and maximization problems.

$$obj_1\,(i) = \min \sum_{i=1}^{n} RMSE_i, AAE_i \tag{9}$$

In Eq. (9), both error values must be minimized by considering total sampling period. If errors are minimized, then the maximization problem can be framed in an automatic mode as follows:

$$obj_2\,(i) = \max \sum_{i=1}^{n} A_i, AC_i \tag{10}$$

There is no need for implementing manual procedures in Eq. (10) as the accuracy of prediction will reach maximum percentage by comparing the true and false values.

RESULTS AND DISCUSSIONS

The experiment is carried out in real time on the Google Collaborator platform with an NVIDIA GeForce graphics card.

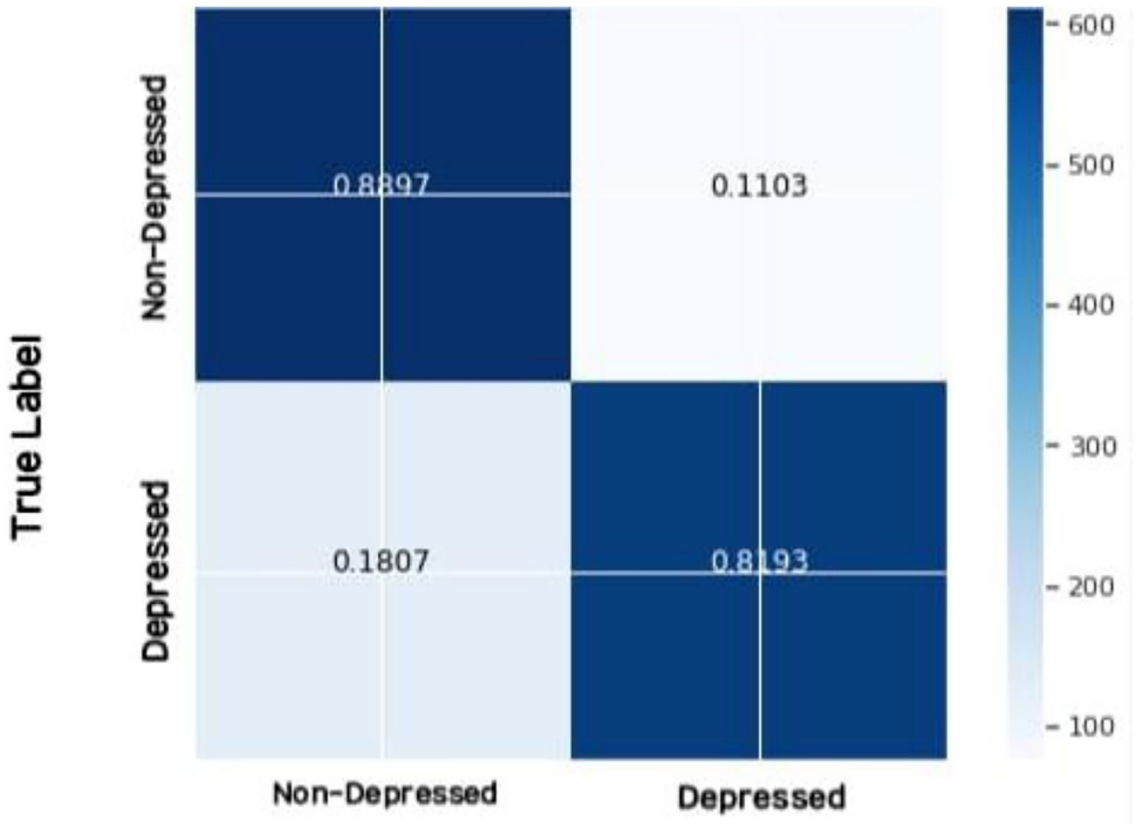


FIGURE 7 | Confusion matrix for depressed and non-depressed classes of the proposed model.

The dataset consisted over 1,500 photos, which were separated into the following three groups in a ratio of 80:10:10: Train set, test set, and validation set. The samples are further classified into eight groups, with scores ranging from 0 to 13 for anxiety and stress, 14 to 19 for mild depression/anxiety, 20–28 for moderate depression/anxiety, and 29–63 for severe depression/anxiety. The training model also includes eight convolutional layers for feature extraction from handwriting and drawing samples as well as a batch activation layer and a batch normalization layer. With modifications in the training dataset, the layered architecture remains the same in each fold. The non-linearity functions in the model are achieved using the Rectified linear unit activation function. The batch normalization procedure is used to normalize the activation from the last layer. The mean value is set to 0 and the deviation to 1 after normalization. The Max pooling approach reduces the dimensionality of the pooling layer. The first, second, and third columns, respectively, show the layer type, input image shape, and number of parameters. While training the model, the output shape of the input image is 64×64 pixels with batch size of none, indicating the input image's dimensionality.

After completing cross-validation on the dataset, eight convolutional layers are blended using an ensemble-based learning process. The suggested ensemble-based model is evaluated on a testing dataset after it has been trained and validated. The convolutional neural filter extracts the features from the handwritten sample photos layer by layer, as shown in the architectural overview. Over a 64×64 image, there are a total of eight 3×3 filters applied. First, the curves and edges are retrieved by the first convolutional layer while the higher features are extracted later by the last convolutional layer. In addition, a BDI-II score column has been added to each of the files for model training. This indicates that features are distinct from one another. In addition, the impact of each feature on the score prediction target variable is investigated. The BDI-II scale is predicted for the tested participants using a random forest regressor with 40 estimators in comparison to the proposed model.

Furthermore, the following hypothesis is considered for analyzing the model's accuracy: A person with a depression scale value is depressed and otherwise not depressed. A binary classification column is added, with the depressed person being labeled as 1. The model's performance is validated by looking for commonalities in categorization and prediction using the participants' labels. Furthermore, a manual and device study are used to make simultaneous observations. If the question was marked "yes" in the recorded session for the domain and the model also gave a higher precision for the same class, it was considered a positive response. Any disagreement or dispute

over classification is met with a negative response. Deep neural networks are used to extract information from handwriting and artwork samples. In addition, to merge these different retrieved feature sets, a fusion method is used. After the features are extracted, a fusion approach is used to combine feature vectors from various modalities in a sequential order. The findings of the comparative analysis will recommend the best performing algorithm for obtaining correct results. Following the extraction of features from handwriting and drawing samples, depression patients are accurately categorized using several ML algorithms, with comparative analysis based on the results.

The SVM algorithm has an accuracy of 82.54%, random forest algorithm has an accuracy of 88.97%, Naive Bayes algorithm has an accuracy of 78.21%, *k*-nearest neighbor algorithm has an accuracy of 71.36%, and logistic regression algorithm has an accuracy of 64.52%, according to **Figure 5** and **Table 2**. With an accuracy of 88.23% and precision of 87.46%, the random forest method was judged to be the best performing algorithm.

As a result, utilizing handwriting and drawing examples to combine their properties to forecast depression scale is an effective strategy. **Table 3** and **Figure 6** show the BDI-II scale for depressive patients utilizing various ML algorithms, with a comparison analysis based on the results.

The RMSE for the SVM algorithm is 9.8548, random forest algorithm is 8.6541, Naive Bayes algorithm is 10.2254, *k*-nearest neighbor algorithm is 10.4141, and logistic regression algorithm is 12.6656, according to comparative analysis. With a mean average error of 8.4235 and a mean error of 8.5696, the random forest method was deemed to be the top performing algorithm.

A confusion matrix is used to visualize the performance of the prediction or classification model. A testing dataset for the known true values is used to generate the table. **Figure 7** depicts the confusion matrix for the depressed and non-depressed classes.

Convergence Characteristics

The depth of integration for an algorithm defines exact coupling strength in real time implementation cases therefore it is necessary to plot convergence of CNN using iteration periods. If a fast convergence rate is attained, then the high benefits can be assured and in turn the sample that is drawn for predicting depression can be identified quickly. In addition, the system time for executing the current state can also be defined using the same characteristic model and is simulated in **Figure 8**. From **Figure 8**, it can be perceived that number of epochs is varied from 10 to 100 and for each period time of convergence is plotted. In all iteration periods, the proposed method using CNN performs well as compared to existing system (32). It can be proved when the iteration reaches 50th period and during this case, the convergence is achieved at 5.77 s whereas the existing model ranges at 8.7 s and at 80th period convergence is attained.

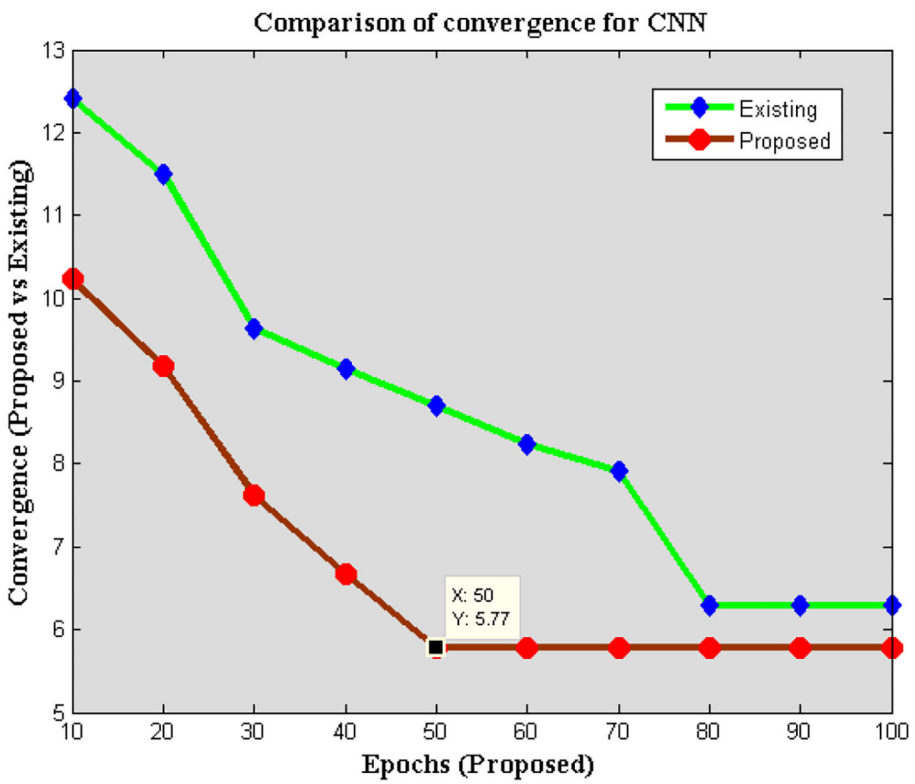


FIGURE 8 | Comparison of convergence characteristics.

This proves that the proposed method is performed at much higher rate for more than 60% in terms of the faster parametric calculations.

CONCLUSIONS

The architecture for predicting the BDI-II scale and diagnosing depression using handwriting and drawing samples was devised and implemented in this study. The model starts with handwriting and drawing samples as input images, which it then runs through a CNN-based classifier. A cloud-based database was created for this system by collecting handwriting text and drawing samples from individuals. Then they were given a task of drawing a pentagon, a house, a circle, and a clock, as well as writing a sentence in cursive. Furthermore, cross-validation in the dataset is performed with many features and were studied from the following handwriting samples: Reference line, angle, distance, control, restraint, arrangement, separation, and sectors, which distinguishes a handwriting sample from other samples. The generated dataset is pre-processed by rotating, scaling, and translating it in order. The photos are resized and normalized to be compatible with the system. The same dataset is utilized to train the deep neural network model, which involves extracting the features from the input image using the neural network architecture and results in BDI-II scale prediction. A comparison of several ML techniques is also performed, and it is discovered that all parametric values are significantly higher than conventional models. In the future, the same approach might be used with high-resolution photos, allowing all flaws to be identified at an early stage of the transformation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary files, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethics review and approval/written informed consent was not required as per local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

PK and HM: conceptualization and methodology. PK, HM, DS, and H-NL: software. SS: data curation. SS and HM: writing—original draft preparation. HA: visualization and investigation. DS and H-NL: supervision, validation, and writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MSIP) (NRF-2021R1A2B5B03002118). This research was supported by the Ministry of Science and ICT (MSIT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-0-01835) supervised by the IITP (Institute of Information and Communications Technology—Planning and Evaluation).

REFERENCES

1. Hemlata H, Manoj M, Kumar S. Personality Detection using Handwriting Analysis Review. In *The Seventh International Conference on Advances in Computing, Electronics and Communication (ACEC).* (2018). p. 85–9. doi: 10.15224/978-1-63248-157-3-33
2. Najafi T, Fomani BA, Shahbahrami A. *Anxiety and Depression Detection using Statistical Features* (2019)
3. Parameswaran NS, Venkataraman D, A. computer vision based image processing system for depression detection among students for counseling. *Indones J Electr Eng Comput Sci.* (2019) 14:503–12. doi: 10.11591/ijeecs.v14.i1.pp503-512
4. Cong Q, Feng Z, Li F, Xiang Y, Rao G, Tao C. X-A-BiLSTM: A Deep Learning Approach for Depression Detection in Imbalanced Data. *Proc - 2018 IEEE Int Conf Bioinforma Biomed BIBM 2018.* (2019) 1624–7. doi: 10.1109/BIBM.2018.8621230
5. Likforman-Sulem L, Esposito A, Faundez-Zanuy M, Clemencon S, Cordasco G, EMOTHAW. A novel database for emotional state recognition from handwriting and drawing. *IEEE Trans Human-Machine Syst.* (2017) 47:273–84. doi: 10.1109/THMS.2016.2635441
6. He L, Cao C. Automated depression analysis using convolutional neural networks from speech. *J Biomed Inform.* (2018) 83:103–11. doi: 10.1016/j.jbi.2018.05.007
7. Mahato S, Goyal N, Ram D, Paul S. Detection of depression and scaling of severity using six channel EEG Data. *J Med Syst.* (2020) 44:118. doi: 10.1007/s10916-020-01573-y
8. Nitti M, Murroni M, Fadda M, Atzori L. Exploiting social internet of things features in cognitive radio. *IEEE Access.* (2016) 4:9204–12. doi: 10.1109/ACCESS.2016.2645979
9. Ashraf A, Gunawan TS, Rahman FDA, Kartiwi M, Ismail N, Ulfiah. A summarization of the visual depression databases for depression detection. *Proc - 2020 6th Int Conf Wirel Telemat ICWT 2020.* (2020) doi: 10.1109/ICWT50448.2020.9243625
10. Niedzwiedz CL, Knifton L, Robb KA, Katikireddi SV, Smith DJ. Depression and anxiety among people living with and beyond cancer: A growing clinical and research priority. *BMC Cancer.* (2019) 19:1–8. doi: 10.1186/s12885-019-6181-4
11. Li Y, Xu Y, Xia M, Zhang T, Wang J, Liu X, et al. Eye movement indices in the study of depressive disorder. *Shanghai Arch Psychiatry.* (2016) 28:326–34. doi: 10.11919/j.issn.1002-0829.216078
12. Kumar R, Nagar SK, Shrivastava A. Depression detection using stacked autoencoder from facial features and NLP. *Ijosthe.* (2020) 7:1–7. doi: 10.24113/ojssports.v7i1.115
13. Raut P. Depression Detection using BDI, Speech Recognition and Facial Recognition. *Int J Res Appl Sci Eng Technol.* (2018) 6:347–51. doi: 10.22214/ijraset.2018.4062
14. Zhu J, Wang Z, Gong T, Zeng S, Li X, Hu B, et al. An Improved classification model for depression detection using EEG and eye tracking data. *IEEE Trans Nanobioscience.* (2020) 19:527–37. doi: 10.1109/TNB.2020.2990690
15. Karplus A. Machine Learning Algorithms for Cancer Diagnosis. *St Cruz Cty Sci Fair 2012* (2012) 1–19.

16. Deshpande M, Rao V. Depression detection using emotion artificial intelligence. *Proc Int Conf Intell Sustain Syst ICISS 2017.* (2018) 858–862. doi: 10.1109/ISS1.2017.8389299
17. Moshe I, Terhorst Y, Opoku Asare K, Sander LB, Ferreira D, Baumeister H, et al. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Front Psychiatry.* (2021) 12:1–12. doi: 10.3389/fpsyt.2021.625247
18. Guo W, Yang H, Liu Z, Xu Y, Hu B. Deep Neural Networks for Depression Recognition Based on 2D and 3D Facial Expressions Under Emotional Stimulus Tasks. *Front Neurosci.* (2021) 15:1–19. doi: 10.3389/fnins.2021.609760
19. Kshirsagar P, Balakrishnan N, Yadav AD. Modelling of optimised neural network for classification and prediction of benchmark datasets. *Comput Methods Biomech Biomed Eng Imaging Vis.* (2020) 8:426–35. doi: 10.1080/21681163.2019.1711457
20. Kshirsagar PR, Manoharan H, Al-Turjman F, Kumar K. Design and Testing of Automated Smoke Monitoring Sensors in Vehicles. *IEEE Sens J.* (2020). doi: 10.1109/JSEN.2020.3044604
21. Manoharan H, Abdul Haleem SL, Shitharth S, Kshirsagar PR, Tirth V, Thangamani M, et al. machine learning algorithm for classification of mental tasks. *Comput Electr Eng.* (2022) 99:107785. doi: 10.1016/j.compeleceng.2022.107785
22. Manoharan H, Teekaraman Y, Kshirsagar PR, Sundaramurthy S, Manoharan A. Examining the effect of aquaculture using sensor-based technology with machine learning algorithm. *Aquac Res.* (2020) 51:4748–58. doi: 10.1111/are.14821
23. Shitharth S, Meshram P, Kshirsagar PR, Manoharan H, Tirth V, Sundramurthy VP. Impact of big data analysis on nanosensors for applied sciences using neural networks. *J Nanomater.* (2021) 2021:1–9. doi: 10.1155/2021/4927607
24. Kshirsagar PR, Manoharan H, Tirth V, Naved M, Siddiqui AT, Sharma AK. Automation monitoring with sensors for detecting covid using backpropagation algorithm. *KSII Trans Internet Inf Syst.* (2021) 15:2414–33. doi: 10.3837/tiis.2021.07.007
25. Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ, et al. metaheuristic optimization approach for energy efficiency in the IoT networks. *Softw - Pract Exp.* (2021) 51:2558–71. doi: 10.1002/spe.2797
26. Cao X zhen, Gadekallu TR. Construction of Sports Safety Information Mining Platform Based on Multimedia Data Sharing Technology. *Mob Networks Appl.* (2022) 1–10. doi: 10.1007/s11036-021-01884-5
27. Bernal J. *Centre de Visió per Computador Bellaterra, Catalonia (Spain)* (2016).
28. Devi BT, Shitharth S. Multiple Face Detection Using Haar - AdaBoosting, LBP-AdaBoosting and Neural Networks. *IOP Conf Ser Mater Sci Eng.* (2021) 1042:012017. doi: 10.1088/1757-899X/1042/1/012017
29. Gadamsetty S, Ch R, Ch A, Iwendi C, Gadekallu TR. Hash-based deep learning approach for remote sensing satellite imagery detection. *Water.* (2022) 14:707. doi: 10.3390/w14050707
30. Amanat A, Rizwan M, Javed AR, Abdelhaq M, Alsaqour R, Pandya S, et al. Deep learning for depression detection from textual data. *Electron.* (2022) 11:1–13. doi: 10.3390/electronics11050676
31. Chung J, Teo J. Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Appl Comput Intell Soft Comput.* (2022) 2022:1–22. doi: 10.1155/2022/9970363
32. Kour H, Gupta MK. An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. *Multimed Tools Appl.* (2022). doi: 10.1007/s11042-022-12648-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kshirsagar, Manoharan, Selvarajan, Alterazi, Singh and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fpubh.2022.919049

Suggestions on Relieving Physical Anxiety of Medical Workers and Improving Physical and Mental Health Under the COVID-19 Epidemic—A Case Study of Meizhou City

Zhi-Cheng Lu[1], *Chien-Hsing Tseng*[2], *Hsiao-Hsien Lin*[1,3], *Yuan-Shing Perng*[4] and *Yi-Han Tseng*[4,5]*

[1] School of Physical Education, Jiaying University, Meizhou, China, [2] Department of Recreation and Sport Management, Shu-Te University, Kaohsiung, Taiwan, [3] Department of Leisure Industry Management, National Chin-Yi University of Technology, Taichung, Taiwan, [4] Department of Forestry, National Chung Hsing University, Taichung, Taiwan, [5] Department of Tourism Leisure and Health Management, Chung Chou University of Science and Technology, Yuanlin City, Taiwan

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Sravan Kumar Reddy,
R G M College of Engineering and Technology, India
D. S Bhupal Naik,
Vignan's Foundation for Science, Technology and Research, India

***Correspondence:**
Yi-Han Tseng
yihjyu@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 13 April 2022
Accepted: 16 May 2022
Published: 17 June 2022

Citation:
Lu Z-C, Tseng C-H, Lin H-H, Perng Y-S and Tseng Y-H (2022) Suggestions on Relieving Physical Anxiety of Medical Workers and Improving Physical and Mental Health Under the COVID-19 Epidemic—A Case Study of Meizhou City. Front. Public Health 10:919049. doi: 10.3389/fpubh.2022.919049

The study examined the effects of swimming pools on healthcare professionals' willingness to engage in recreational activities, physical anxiety, and physical and mental well-being in the context of COVID-19. The research adopted the mixed research method, used SPSS 26.0 statistical software to test the reliability of the questionnaire, and then collected 840 valid questionnaires; first analyzed the data with basic statistics, *t*-test, ANOVA, and PPMCC test methods, and then used the interview method to collect expert opinions. A multi-check approach assembled all data and discussions. The study found that the use of personnel dynamic tracking systems or measures, combined with sodium hypochlorite and repeated filtration to stabilize water quality, could maintain the confidence of most medical workers in the swimming pool sports environment for epidemic prevention and avoid violations. The government could formulate safety prevention and control mechanisms in traffic and establish appropriate traffic routes. Next, formulated a prescription for swimming or other physical activity mechanisms for men aged 31–50 and redesigned measures for medical staff over 51 years old to have tense head issues and physical fatigue, promote blood circulation and improve sleep quality. This will promote the purpose of relieving stress and regulating the physical and mental health of medical staff after engaging in swimming.

Keywords: SARS-CoV-2, healthcare professionals, environment risk, body anxiety, leisure satisfaction, physical and mental health

INTRODUCTION

Since the outbreak of COVID-19 in December 2019, humanity has been facing one of the most difficult times in recent global epidemic cases and is still not free from the threat (1). In recent years, governments around the world have vigorously promoted vaccine management policies, invested a lot of medical resources, and hoped that the popularization of vaccines would increase people's

resistance (2). Governments actively call on the people to wear masks, wash their hands frequently, avoid gatherings and talk closely to avoid being infected by the virus and even spreading the epidemic (3). In recent years, a number of studies have also proposed ways to teach people how to maintain regular exercise to maintain health in the context of COVID-19 (4–9). It is expected that exercise can improve physical fitness, enhance human immunity, relieve the pressure of medical staff, achieve the goal of resisting viruses, and enhance individual physical and mental health (10).

Because SARS-CoV-2 is an enveloped virus, it is highly infectious, hydrophilic, and difficult to eradicate (11). And the virus gene gradually adapts to the human body, and the virus gene is constantly mutated (12), which increases virus transmission efficiency (13). However, the epidemic threat still exists today, and the environment is full of infection risks, which hinders the willingness to go out (14). As a result, people are full of insecurity in the social environment (10, 14), causing people to panic, anxiety, loss and other psychological pressures, affecting their physical health (3, 15). So far, there have been 513 million cases of infection in 199 countries around the world, with 6.23 million deaths, and infected people continue to be found (16). This phenomenon has impacted the medical systems of various countries, seriously the workload of medical staff and exhausting their physical and mental health (17). Therefore, some scholars believe that finding ways to improve the physical and mental stress of medical staff, provide a safe leisure environment, and promote physical and mental health will be a good way to maintain the service quality of medical staff and stabilize national public health management (10, 18). So we think this is an important research topic at present.

Swimming has the functions and characteristics of enhancing cardiopulmonary function, improving physical fitness, and improving physical and mental health (19, 20). Many countries actively build swimming pools to provide a safe and hygienic recreational sports environment, allowing people to continue swimming to maintain health (21, 22). In order to maintain the sanitation and cleanliness of swimming pool water, sodium hypochlorite has been widely used for water disinfection and environmental cleaning for a long time (21). This method has been recognized by the World Health Organization and the National Centers for Disease Control and Prevention and stated that if the effective residual chlorine in the swimming pool can be maintained above 1 ppm, it can effectively inhibit Escherichia coli and Legionella (22, 23), and decompose bacterial nuclei (24). This measure is applied to enveloped virus-type pathogens such as SARS-CoV-2. As long as the concentration of sodium hypochlorite is increased to 0.5% (5,000 ppm), highly resistant pathogens can be effectively inhibited (25–28). In addition, through effective filtration facilities, impurities in swimming pool water are repeatedly filtered to maintain the hygiene of water quality (29, 30), and sodium hypochlorite is used for environmental disinfection (27, 28). Some scholars believe these measures will provide swimming pools with a safe and hygienic exercise environment during the epidemic (25–30). Therefore, we believe that when the surrounding environment of the swimming pool and the quality of the water are guaranteed, it should be possible to create a safe leisure sports environment (31, 32). Therefore, if the medical staff is advised to go swimming in swimming pools, they should be able to obtain regular exercise, enhance cardiopulmonary function, improve physical fitness, reduce anxiety, and stabilize physical and mental health under a low-risk awareness of the exercise environment (20, 21).

However, since the current problem is not under control, there has always been an infection risk in the surrounding environment (15), which seriously affects people's willingness to go out. Therefore, there is a gap between plan and reality (8, 10). Therefore, some scholars suggest that it is necessary to understand the risk perception of the current swimming pool by medical staff engaged in swimming, as well as the changes in personal anxiety, exercise benefits, and physical and mental health feelings after going to exercise (32). However, these changes and experiences take time (33, 34). Through the awareness of environmental risk assessment, people can assess the potential crisis and risk degree of the surrounding environment (35). Through the perception of physical anxiety, medical staff's personal anxiety changes after engaging in swimming can be found (32). Through the perception of leisure satisfaction, we can find the changes in personal leisure satisfaction after medical staff engage in swimming (36). Finally, the physical and mental health perception can be used to understand the improvement of personal physical and mental health after medical staff engages in swimming (37–39). These answers can be corroborated by participants' feelings (34, 40, 41). Therefore, we believe that by discussing environmental risks, physical anxiety, leisure satisfaction will effectively improve the environmental risk perception of medical personnel, reduce anxiety, improve leisure benefits, and achieve the goal of promoting physical and mental health, and physical and mental health perception, we can know whether the physical and mental health of medical staff can be improved after swimming.

Furthermore, after analyzing the literature, we know that although the current research directions on COVID-19 and human health problems are quite extensive (1, 2, 4, 5, 17). However, most of them discussed the physical and mental health pressure and epidemic prevention measures during the epidemic (5, 25–30, 42). Although the impact of participating in tourism activities during the epidemic has also been discussed (33, 36, 43, 44), methods to improve people's health benefits have also been proposed, such as Qigong (6), intermittent fitness exercises (7), mountain climbing (4), and jogging (8). More recently, related impacts of leisure, travel, living and working environment, physical and mental health have also been discussed (36–38), as well as environmental risks, satisfaction, and physical and mental health issues (40, 41, 45). However, the researchers found that swimming was not discussed as a topic, and the environmental risks, physical anxiety, leisure satisfaction, and physical and mental health perceptions of healthcare workers engaged in swimming in SARS-CoV-2 settings.

Based on the above descriptions, we believe that taking the epidemic period as the timeline, and collecting the opinions of medical staff who continue to swim during the epidemic from the aspects of environmental risk, physical anxiety, leisure satisfaction, and physical and mental health cognition, the truth

will be obtained. These answers will help the government, medical institutions, or medical staff finds a safe exercise environment and regular exercise methods, provide effective exercise, improve resistance, maintain medical quality, and stabilize public health orders during the COVID-19 epidemic.

LITERATURE REVIEW

Leisure Environment Risk

Risk refers to the potential factors that people may develop adversely, personally, or financially (35). Sports environmental risk is mainly to understand the unpredictable but potentially far-reaching potential problems that arise when people or society face the environmental hazards of the epidemic (46). Risk causes are often at odds with everyday knowledge, even conflicting or unpredictable (47). The individual's reaction to risk is the environmental risk perception (48).

Establishing a risk requires three elements: uncertainty, the possibility of loss, and future attributes (49). Individuals feel uncertain about future outcomes, and the outcome may have a negative impact, and the higher the uncertainty, the greater the risk (50). Some scholars believe that predicting risks can usually be discussed in terms of behavior, facility and environmental safety, potential risks, economic losses, and increased social costs (48–51).

Therefore, the researchers believe that the current situation of swimming pools should be referred to as environmental risk prediction indicators. From the aspects of exercise mode, equipment safety, movement fluency, potential risks, happiness, and social costs, it is possible to understand the environmental risk perception of the public when exercising in the swimming pool.

Physical Anxiety

Anxiety refers to an individual's state of excessive, widespread, persistent, and perceived uncontrollable worry about many events or activities, almost most of the time, for at least the past 6 months (52). Anxiety assessment mainly aims to understand people's current psychological mood and distress from different aspects and provide suggestions to improve the quality of life (53). Body anxiety refers to an individual's negative subjective perceptions and feelings that stimulate the body with the environment or time (54).

The establishment of anxiety requires a pathogenic context. The operation process of a specific system, such as cognition related to information processing, emotional stimulus processing, coping, and psychophysiological aspects, are key factors (55). If the individual is worried about the phenomenon or event that continues to occur, or even a catastrophic phenomenon occurs, it may cause psychological trauma (56). Some scholars believe that physical anxiety can be assessed from the aspects of physiological behavior, psychological status, behavioral performance, and effectiveness judgment (57).

Therefore, researchers believe that the answer should be obtained based on the physical anxiety state that the research subjects may face and then referring to relevant literature on physical anxiety, psychological behavior settings (from motor behavior anxiety, tension), physiological behavior (from limb stiffness, muscle tension, stomach tension, rapid heartbeat, and other behavior), and effectiveness judgments (from the affirmation of athletic performance, achievement of expected goals, exceeding levels, and zero mistakes) to predict the state of physical anxiety.

Leisure Satisfaction

Leisure satisfaction refers to the positive perceptions of individuals engaged in leisure activities after comparing their perceived leisure experience with the actual environment (58). To put it simply, people judge according to the current leisure experience whether the acquired individual needs and satisfaction meet the expected goals, which is leisure satisfaction (43).

Leisure satisfaction uses personal activity experience to compare prior experience, personal expectations, or expectations (59). When the actual leisure environment or activity content conforms to the individual's expectations, a satisfied attitude is generated in the heart (58, 59). The higher the satisfaction, the higher the positive energy cognition obtained (60). The more satisfaction one can obtain from personal psychological needs, the more satisfied with the current life and feelings, and the more willing to revisit (58). Some scholars believe that although the influencing factors of leisure satisfaction can be divided into internal and external interference (61), it can be discussed in terms of decision-making measures, environment, facilities, and personal leisure benefits (59–61).

Therefore, researchers believe that the answer should be obtained with reference to the literature on leisure satisfaction and according to the problems that the research topic may encounter, such as decision-making measures that maintain a safe distance, people wearing masks, keeping track of personnel movements, and the environment and facilities being discussed in terms of personal hygiene of swimmers, clean water, and clean facility environment.

Cognition of Physical and Mental Health

Physical and mental health refers to people's physical, psychological, and social aspects to achieve a state of well-being (59). It can present the actual situation through scientific testing evidence such as self-assessment (58), which belongs to the analysis method of self-perception assessment (60). When the public perceives a higher health risk, the power to influence an individual's final behavioral decision will be greater (60).

Physical and mental health cognition mainly investigates the current state of individual physical and mental health and presents the impact of the current environment on people (62). Some scholars believe that physical and mental health can be divided into four levels: mental, spiritual, physical, and negative attitudes (60, 61). And people can get answers from 16 questions such as learning, fear, performance, lack of interest in things, lack of time, headache, fatigue, back pain, insomnia, stomach pain, overeating, muscle stiffness, tantrums, loss, burnout, depression, and suicide (60–63).

Therefore, the researchers believe that the answers should be obtained by referring to those mentioned above in physical and

mental health-related literature, according to the psychological, spiritual, physical, and negative attitudes and other aspects, and with the evaluation based on 16 indicators, such as learning, fear, performance, lack of interest in things, and lack of time.

METHODS

Framework

The study first collected relevant literature, analyzed the influence of swimming pool disinfection and filtration management measures on epidemic prevention and control, and referred to literature on environmental risks, physical anxiety, leisure satisfaction, and physical and mental health cognition. Took the epidemic period as the timeline, targeted those who continue to engage in swimming, and based on the literature on environmental exercise risk (46–51), physical anxiety (52–57), leisure satisfaction (58–61), and physical and mental health cognition (58–63), the study was summarized and analyzed. Looking forward to knowing whether swimming sports and venues would have the effect of providing people with a healthy and safe leisure environment, reducing personal anxiety, and promoting physical and mental health. The research structure is shown in **Figure 1**.

Process and Analysis

At present, the research direction on COVID-19 and human health issues has been quite broad (1–30). However, there has been no discussion about the environmental risks, physical anxiety, leisure satisfaction, and physical and mental health cognition of people engaging in swimming in the context of SARS-CoV-2 with swimming as the theme. Therefore, mixed research methods were used through the feelings of actual participants (33, 34, 40, 44), quantitative research methods were used to supplement the research breadth (64, 65), and qualitative research methods were used to increase research depth (66). Then the research would be discussed in a multi-check method (67–69), which could make up for the shortcomings of research methods or insufficient theory (70).

Therefore, the researchers first referenced literature on environmental risk (46–51), physical anxiety (52–57), leisure satisfaction (58–60), and physical and mental health perceptions (60–63), and then developed a questionnaire tool. First, SPSS 26.0 statistical software was used to confirm the framework of the formal questionnaire tool with basic statistical verification and EFA verification method, and then three experts with expertise in public health, leisure sports, and decision analysis were invited to check the content validity to confirm the formal questionnaire. In November 2021, samples were collected by interest sampling method and snowball sampling method, and 840 questionnaires were finally obtained.

The data were analyzed by basic statistics, *t*-test, ANOVA, and PPMCC test. Then, the semi-structured interview method was used to collect the opinions of officials, experts, and the public according to the results of the questionnaire analysis. Next, all the data in a rigorous was collected in an orderly and logical manner and then valuable information through summarizing, organizing, and organizing methods were summarized (69). Finally, the research was carried out through the multi-check method in a multi-data and multi-perspective way (67, 68, 71).

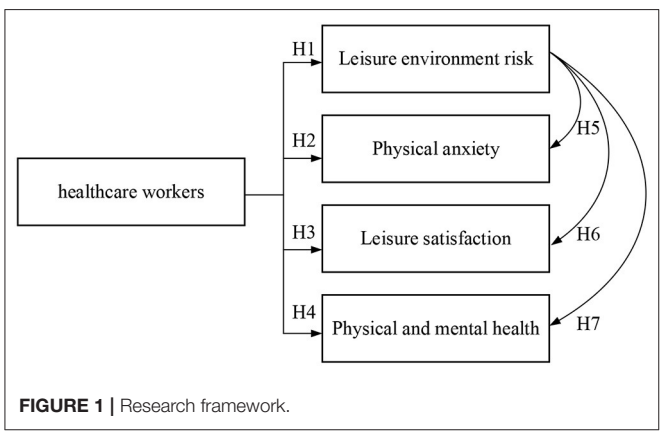


FIGURE 1 | Research framework.

Hypothesis

Due to the impact of the potential risk of epidemic infection around the world, people's willingness to go out for leisure and activities has been declined (13). As many studies had confirmed that swimming pool disinfection and filtration management measures had a certain bactericidal effect on the SARS-CoV virus (24–26), the conditions were sufficient to create a healthy and safe leisure sports environment (29, 30). These were beneficial for people to concentrate and relax their bodies during leisure sports (33, 34), effectively obtain the effect of leisure sports (19, 72), and achieve the purpose of promoting physical and mental health (19, 72). Therefore, the researchers put forward five hypotheses based on the above inferences and research framework.

Hypothesis 1. Assume that the risk perceptions of the exercise environment for swimming in swimming pools are consistent.
Hypothesis 2. The body anxiety perceptions of swimming in the pool are consistent.
Hypothesis 3. The perception of leisure satisfaction of swimming in swimming pools is consistent.
Hypothesis 4. The physical and mental health perceptions of swimming in the pool are consistent.
Hypothesis 5. Environmental risk has a positive and significant effect on physical anxiety, leisure satisfaction, and physical and mental health.

Research Tools and Analysis

The study used a mixed approach, with references to environmental risk (46–51), physical anxiety (52–57), leisure satisfaction (58–61), and perceptions of physical and mental health (37, 58–61), and then prepared a preliminary questionnaire. The questionnaire was divided into background information and variables. The background was gender (male, female) and age (under 30, 31–40, 41–50, over 51). Other variables were divided into environmental risk, physical anxiety, leisure satisfaction, and physical and mental health. There were 42 questions in total and were designed with a Likert-style

5-point scale (1 point meant very dissatisfied, 5 points meant very satisfied). After the preliminary questionnaire was edited, according to the Kaiser-Meyer-Olkin (KMO) > 0.06 in the test results, and the p-value in the Bartlett test was <0.01 ($p < 0.01$) (73), $\alpha > 0.60$. The questionnaire met the standard of good reliability issues (74), and the content of the questionnaire applicable to the subject of this research was reserved for continuous research and analysis.

According to references, edited environmental risk design six questions (47–51). The analysis results showed that the KMO was 0.837, the approximate $\chi 2$ value of Bartlett was 970.578, the df was 6, and the significance was $p < 0.001$, which was suitable for factor analysis. The total variance explained by the scale was 67.414%. After factor analysis, good and reliable topics were retained. The alpha coefficient of the total scale was 0.891. After deleting the inappropriate topics, they were four topics reserved at the end, "modifying body movements," "exercise methods will not cause infection risks," "good mood," and "no increase in social costs." Each topic scored 0.855 to 0.862. According to the above analysis results, it could be known that this questionnaire had good reliability.

Referring to (52–57) literature, 10 questions were designed to edit body anxiety cognition. The analysis results showed that the KMO was 0.683, the approximate $\chi 2$ value of Bartlett was 2,634.081, the df was 28, and the significance was $p < 0.001$, which was suitable for factor analysis. The variance explained by the scale was 38.131 and 30.148%, and the total explained variance was 68.28%. After factor analysis, good and reliable issues were retained and named psychological and physical (5 questions) and achievement performance (3 questions). The alpha coefficient of the total scale was 0.851. After deleting the inappropriate items, they were reserved at the end. There were 8 questions in total, including anxiety, tension, physical incoordination, muscle relaxation, stomach pain, performance satisfaction, achievement of goals, and surpassing level. Each question was 0.809–0.849. According to the above analysis results, it could be known that this questionnaire had good reliability.

Referring to the literature (58–61), a total of nine questions were designed to edit the cognition of leisure satisfaction. The analysis results showed that the KMO was 0.683, the approximate $\chi 2$ value of Bartlett was 2,634.081, the df was 28, and the significance was $p < 0.001$, which was suitable for factor analysis. The variance explained by the scale was 38.131 and 30.148%, and the total explained variance was 68.28%. After factor analysis, good and reliable issues were retained and named as epidemic prevention measures (3 questions), water quality and environment (3 questions), and leisure effects (2 questions). The alpha coefficient of the total scale was 0.908. After deleting the inappropriate topics, eight questions were finally retained 8 questions in total, including wearing masks, epidemic prevention distance, source of swimmers, personal epidemic prevention and hygiene literacy, clean water quality, clean facilities, environment, exercise ability, and completed actions. Each question scored 0.894–0.897. According to the above analysis results, it could be known that this questionnaire had good reliability.

Referring to the literature (62–65), a total of 17 questions were designed to edit physical and mental health cognition. The analysis results showed that the KMO was 0.756, the approximate $\chi 2$ value of Bartlett was 8,613.85, the df was 136, and the significance was $p < 0.001$, which was suitable for factor analysis. The variance explained by the scale was 38.131 and 30.148%, and the total explained variance was 68.28%. After factor analysis, good and reliable issues were reserved and named as psychological feelings (5 questions), mental state (5 questions), physical conditions (3 questions), and negative thoughts (4 questions). The alpha coefficient of the total scale was 0.908. After deleting the inappropriate topics, the remaining topics were facing environmental pressure, feeling scared, unsatisfactory performance, not interested in life, unable to make good use of time, head pressure, fatigue, pain sensitivity, back tightness, insomnia, stomach pain, eating disorders, body tension, easy tantrums, worthless, life burnout, and suicidal escapism, a total of 17 questions. Each question scored 0.894–0.897. According to the above analysis results, it could be known that this questionnaire had good reliability.

According to the above analysis results, there were 37 questions left in the background disguised, environmental risk (4 questions), physical anxiety (8 questions), leisure satisfaction (8 questions), and physical and mental health (17 questions). Finally, three experts were asked to check the content validity, confirm the content of the questionnaire, and confirm the completion of the formal questionnaire, and then investigation and analysis were conducted.

Research Scope and Object

Meizhou City is located in the mountainous area of Guangdong Province. Since the outbreak of the epidemic in 2019, the epidemic control mechanism has been continuously adjusted. Although there are still sporadic outbreaks of overall infection cases in various provinces, a number of strict control measures have been carried out on the epidemic situation in the case areas. So far, they have been effectively controlled, and no large-scale infection incidents have occurred (75). In addition, the local latitude is high, the climate change is small, and the people's willingness to invest in swimming is relatively high.

Therefore, the research team limited the sampling to medical workers who were still engaged in swimming activities during the epidemic. Taking Meizhou City as the core of the research and investigation scope, the interest sampling method was used to collect the basic sample size. At the same time, the snowball sampling method was used to invite people who had agreed to be surveyed to assist in recommending, and other samples were collected through the online questionnaire platform, which was radiated and diffused outwards, and finally, 840 samples were obtained.

In addition, the interview process invited six experts in decision analysis, public health, and leisure sports management, as well as long-term swimmers, the general public, and industry players, to obtain the interviewee's consent to the interview by using the video system and telephone calls. Then the semi-structured interview method was used to conduct the survey, and opinions were put forward on the questionnaire analysis and the results. The interview subjects and interview topics are described in **Table 1**.

TABLE 1 | Respondent's background information and an overview of the interview outline.

Identity	Gender	Residence and sports time/years of work experience	Identity	Gender	Residence and sports time/years of work experience
Swimmer	Male	31	Professor	Female	15
Swimmer	Female	41	Entrepreneur	Female	30
Professor	Male	20	Entrepreneur	Male	10
Professor	Male	16	People	Male	40
Construct	**Issues**				
Impact of tourism development	1. Do you agree that during the exercise in the swimming pool, the sports environment still feels the threat of the epidemic? What parts? What is the reason? 2. Do you agree that physical and mental health will be threatened during exercise in the swimming pool? Which parts? Why? 3. Do you agree that leisure satisfaction is affected by exercising in the swimming pool? Which parts? Why? 4. Do you agree that your physical and mental health will be affected by exercising in the swimming pool? Which parts? Why?				

Ethical Considerations

This study was conducted using a mixed-type study, with Meizhou City as the scope. The standard for collecting questionnaire samples is local people with medical work status who have continued to swim since the beginning of the epidemic. Investigators produced a brief description of the swimming pool safety research report, briefed respondents or asked respondents to read the content during the sampling period, and then began the survey. Respondents, therefore, had an understanding of the research topic and had relevant professional competence or personal experience. Therefore, all respondents agreed to and understood the purpose of this research and then cooperated with the presentation of anonymous data and under the condition of knowledge, recorded and collected data, and agreed to cooperate with the provision of relevant data for data collection. The research design and data collection process were carried out under the standards of fairness, impartiality, and openness. Therefore, the researchers believed that the research conformed to ethical standards and did not need to apply for ethical certification (75, 76).

However, during the research and investigation period, there were still some cases of epidemic transmission in China. In order to avoid the risk of infection, the research and investigation team only took the local area of Meizhou City as the on-site distribution scope. Considering the limitations of resources such as human resources, material resources, and funds of the team, as well as factors such as the time node of swimming participation and the location of swimming pools for assisting pushers, the above limitations might affect the total number of samples and analysis results. Therefore, if the results of this study were insufficient, they would be listed as recommendations for follow-up research investigations in the hope that follow-up researchers would strengthen them.

ANALYSIS AND DISCUSSIONS

The study used basic statistics, *t*-test, ANOVA, and PPMCC test methods to analyze the related issues among background information, environmental risks, physical anxiety, leisure satisfaction, physical and mental health cognition, and variables in 840 samples.

The analysis revealed that among the 840 samples, 276 (32.2%) were male and 580 (67.8%) were female. Among the age groups, 304 (35.5%) were under the age of 30, 376 (43.5%) were 31–40 years old, 152 (17.8%) were 41–50 years old, and 28 (3.3%) were over 51 years old. It could be seen that most of the samples tested were women, and the least were men; people under the age of 40 were the majority, and those over the age of 51 were the least.

Analysis of People's Sports Environment Risk, Physical Anxiety, Leisure Satisfaction, Physical and Mental Health Cognition

The analysis used basic statistics, *t*-test, and ANOVA to analyze the environmental risks, physical anxiety, leisure satisfaction, and physical and mental health cognition of people swimming in swimming pools under the epidemic and verify Hypotheses 1–4. The analysis found that the modified action (3.95) was the highest in the risk environment, and the exercise mode did not have the lowest risk of infection (3.77). In physical anxiety, tension (3.52) and performance satisfaction (3.3) were the highest, muscle relaxation (3.0) and goal achievement (3.14) were the lowest. Leisure satisfaction was the highest in terms of distance from epidemic prevention (3.61), clean water (3.6), exercise ability (3.58) while wearing a mask (3.55), clean facilities (3.55), and completing actions (3.55) were the lowest. Physical and mental health was the highest with the inability to use time well (2.79), insomnia (2.87), body tension (2.79), and easy to lose temper (2.58). Dissatisfaction with performance (2.65), head pressure (2.41), stomach pain (2.48), and no value (2.33) were the lowest. The above analysis results confirmed that Hypotheses 1–4 were invalid.

In the follow-up analysis, gender was insignificant in environmental risk and leisure satisfaction ($p > 0.01$). In terms of physical anxiety, tension, limb incoordination, achievement of goals, and transcendence were significant ($p < 0.01$), and the others were not significant. Men felt high in tension and

TABLE 2 | Analysis of people's sports environment risk, physical anxiety, leisure satisfaction, physical and mental health cognition.

			M	Rank	Gender					Age					
					Male		Female		*p*	Under 30 (M)	31–40 (M)	41–50 (M)	Over 51 (M)	*p*	*Post-hoc*
					M	SD	M	SD							
Sports environment risk		Modify body movements	3.95	1	3.93	0.731	3.96	0.692	0.601	4.01	3.90	3.95	4.00	0.337	NA
		No risk of infection	3.77	4	3.68	0.837	3.82	0.765	0.084	3.84	3.73	3.76	3.64	0.542	NA
		Good mood	3.88	2	3.83	0.779	3.91	0.682	0.066	3.97	3.85	3.76	4.00	0.002*	NA
		No social cost	3.85	3	3.87	0.733	3.84	0.717	0.828	3.89	3.80	3.89	3.93	0.500	NA
Physical anxiety	Mental and physical	Anxiety	3.37	3	3.67	0.738	3.23	0.902	0.651	3.25	3.35	3.63	3.43	0.094	NA
		Nervous	3.52	1	3.97	0.639	3.30	0.929	0.000*	3.36	3.51	3.68	4.57	0.101	NA
		Limb incoordination	3.48	2	3.71	0.642	3.37	0.997	0.000*	3.29	3.52	3.55	4.57	0.028	NA
		Muscle relaxation	3.00	5	2.81	0.892	3.08	1.029	0.804	3.00	3.25	2.47	2.43	0.071	NA
		Stomachache	3.03	4	2.81	1.084	3.14	1.069	0.209	2.93	3.34	2.50	2.86	0.263	NA
	Achievement performance	Performance satisfaction	3.30	1	3.25	0.772	3.32	0.695	0.168	3.30	3.47	2.71	4.14	0.000*	Under 30, 31–40, over 51 > 41–50
		Achieve goals	3.14	3	3.16	1.048	3.14	0.731	0.000*	3.21	3.33	2.45	3.71	0.000*	31–40 > under 30, 41–50; over 51 > 41–50
		Beyond the level	3.21	2	3.09	1.091	3.27	0.800	0.001*	3.11	3.42	2.82	3.71	0.000*	Under 30, 31–40 > 41–50, over 51; 41–50 > over 51
Leisure satisfaction	Epidemic prevention measures	Wear mask	3.55	3	3.56	1.190	3.54	1.205	0.927	3.48	3.53	3.72	3.57	0.759	NA
		Epidemic prevention distance	3.61	1	3.59	1.182	3.63	1.170	0.872	3.58	3.65	3.58	3.71	0.196	NA
		Source of bathers	3.60	2	3.62	1.166	3.59	1.194	0.940	3.64	3.52	3.71	3.71	0.523	NA
	Water quality	Personal epidemic prevention literacy	3.59	2	3.62	1.161	3.57	1.139	0.978	3.53	3.56	3.67	4.07	0.230	NA
		Water quality	3.60	1	3.63	1.203	3.59	1.237	0.686	3.55	3.56	3.75	3.86	0.201	NA
		Clean facility	3.55	3	3.58	1.213	3.53	1.153	0.406	3.51	3.49	3.71	3.71	0.360	NA
	leisure effect	athletic ability	3.58	1	3.55	1.215	3.59	1.217	0.971	3.57	3.53	3.63	4.00	0.038	NA
		Complete the action	3.55	2	3.60	1.199	3.53	1.123	0.346	3.61	3.44	3.71	3.64	0.878	NA

(Continued)

TABLE 2 | Continued

			M	Rank	Gender					Age					
					Male		Female		*p*	Under 30 (M)	31–40 (M)	41–50 (M)	Over 51 (M)	*p*	*Post-hoc*
					M	SD	M	SD							
Physical and mental health	Psychological feeling	Facing environmental pressure	2.75	2	2.55	1.061	2.85	0.867	0.000*	2.72	2.87	2.34	3.71	0.000*	41–50 < under 30, 31–40, over 51
		Feeling scared	2.67	4	2.51	0.976	2.74	0.902	0.049	2.49	2.86	2.37	3.71	0.000*	31–40 > under 30, 41–50, over 51; over 51 > under 30, 31–40, 41–50
		Unsatisfactory performance	2.65	5	2.64	1.053	2.66	0.866	0.003*	2.42	2.86	2.34	4.14	0.072	
		Not interested in life	2.72	3	2.38	0.803	2.88	0.995	0.848	2.67	2.94	2.32	2.57	0.184	
		Can't make good use of time	2.79	1	2.71	0.906	2.82	0.939	0.126	2.57	2.98	2.50	4.14	0.036	
	Mental condition	Head pressure	2.41	5	2.16	1.089	2.53	0.873	0.005*	2.16	2.87	1.66	3.14	0.010*	Under 30 > 41–50, over 51; 31–40 > under 30, 41–50; over 51 > under 30, 41–50
		Tired	2.44	4	2.61	1.014	2.36	0.862	0.003*	2.29	2.66	2.11	3.00	0.000*	31–40 > under 30, 41–50; over 51; over 51 > under 30, 41–50
		Pain sensitivity	2.58	3	2.51	1.102	2.61	1.047	0.505	2.50	2.77	1.97	4.14	0.000*	Under 30 > 41–50, over 51; 31–40 > 41–50; over 51 > under 30, 31–40, 41–50
		Tight back	2.62	2	2.59	1.150	2.63	1.065	0.083	2.34	2.80	2.37	4.57	0.002*	31–40 > under 30, 31–40, 41–50; over 51 > under 30, 31–40, 41–50
		Insomnia	2.87	1	3.00	1.232	2.81	0.948	0.000*	2.75	2.86	2.84	4.57	0.000*	over 51 > under 30, 31–40, 41–50
	Physical conditions	Stomachache	2.48	3	2.19	0.892	2.62	0.942	0.215	2.36	2.85	1.92	2.00	0.000*	under 30 > 41–50, over 51; 31–40 > under 30, 41–50, over 51
		Eating disorders	2.63	2	2.46	1.141	2.70	1.026	0.001*	2.46	2.92	2.42	1.57	0.000*	31–40 > under 30, 41–50, over 51; under 30 > over 51

(Continued)

TABLE 2 | Continued

		M	Rank	Gender					Age				
				Male		Female		*p*	Under 30 (M)	31–40 (M)	41–50 (M)	Over 51 (M)	*p*	*Post-hoc*
				M	SD	M	SD							
Negative thoughts	Limbs tense	2.79	1	2.70	0.860	2.83	0.936	0.163	2.64	3.02	2.53	2.57	0.530	
	Easy to lose temper	2.58	1	2.17	0.903	2.78	0.827	0.063	2.32	2.88	2.42	2.43	0.080	
	No value	2.33	3	2.12	1.047	2.43	0.870	0.003*	2.17	2.61	2.08	1.57	0.136	
	Life burnout	2.46	2	1.96	0.973	2.70	1.006	0.661	2.49	2.73	1.92	1.57	0.004*	Under 30, 31–40 > 41–50, over 51
	Suicide to escape reality	2.28	4	1.68	0.928	2.56	1.038	0.045	2.22	2.70	1.47	1.57	0.089	

*$P < 0.01$.

limb incoordination, and women felt high in transcendence level. In terms of physical and mental health cognition, facing environmental pressure, performance dissatisfaction, head pressure, fatigue, insomnia, eating disorder, and no value were significant ($p < 0.01$), and the others were not significant. And men had higher feelings of head pressure, fatigue, and insomnia, while women had higher feelings of environmental stress, unsatisfactory performance, eating disorders, and worthlessness.

In addition, subjects of different ages toward the issues in good moods, showing satisfaction, reaching goals, exceeding levels, facing environmental stress, feeling scared, head pressure, fatigue, pain sensitivity, back tightness, insomnia, stomach pain, eating disorders, and life burnout were significant ($p < 0.01$), and others were not. In addition, those under the age of 30 were more sensitive to the performance of physical anxiety, exceeding the level, as well as physical and mental health issues such as head pressure, pain sensitivity, stomach pain, eating disorders, and life burnout. Those aged 31–40 were more sensitive to physical anxiety, performance satisfaction, achievement of goals, and exceeding levels, as well as physical and mental health issues such as fear, pain sensitivity, back tightness, stomach pain, eating disorders, and life burnout. Those aged 41–50 were more sensitive to environmental stress issues in achieving physical and mental health. People over 51 years old were satisfied with their performance in physical anxiety, and their physical and mental health were more sensitive to fear, head pressure, pain sensitivity, back tightness, and insomnia.

It could be seen that most people believed that swimming could modify body movements, but there was still a risk of infection. Although the performance of the exercise was satisfactory, the body was still tense, the muscles could not relax, and the exercise goal could not be achieved. Even if the distance of epidemic prevention and the clean water were reassuring, the exercise ability could be improved. However, problems such as wearing masks by the swimmers and the cleanliness of the facilities remained, and the swimmers could not complete the movements, resulting in a lack of time, insomnia, and physical tension among the people's physical and mental health. These caused different genders to have cognitive differences in physical anxiety and physical and mental health cognition. Different ages also had different views on physical anxiety and physical and mental health cognition. As shown in **Table 2**.

Correlation Analysis of Physical Anxiety, Leisure Satisfaction, and Physical and Mental Health Cognition of Sports Environmental Risks

The PPMCC test method was used to analyze the impact of environmental risks on physical anxiety, leisure satisfaction, and physical and mental health cognition of people who still engage in swimming during the epidemic. The analysis showed that environmental risk had a positive and significant positive effect on physical anxiety (0.299) and leisure satisfaction (0.299), and $p < 0.01$, but had no significant effect on physical and mental health cognition. Among them, environmental risk cognition had the greatest influence on achievement performance (0.298)

TABLE 3 | Correlation analysis of sports environmental risk on physical anxiety, leisure satisfaction, and physical and mental health cognition.

Sports environment risk	Physical anxiety			Leisure satisfaction			
	Overall facet	Mental and physical	Achievement performance	Overall facet	Epidemic prevention measures	Water quality	Leisure effect
Overall facet	0.299**	0.253**	0.298**	0.137**	−0.030	−0.126*	0.370**
Modify body movements	0.013	0.001	0.028	0.023	−0.022	0.059	0.035
No risk of infection	0.036	0.010	0.072	0.000	−0.034	0.054	0.009
Good mood	0.003	−0.020	0.050	0.048	0.043	0.093	0.003
No social cost	−0.017	−0.025	0.000	0.045	−0.006	0.101	0.044

Sports environment risk	Physical and mental health				
	Overall facet	Psychological feeling	Mental condition	Physical conditions	Negative thoughts
Overall facet	−0.030	0.128**	−0.046	−0.089	−0.136**
Modify body movements	−0.009	0.012	−0.009	−0.019	−0.022
No risk of infection	−0.025	−0.003	−0.037	−0.044	−0.013
Good mood	0.034	0.055	0.028	0.023	0.010
No social cost	−0.011	0.008	−0.010	−0.032	−0.015

*$p < 0.05$; **$p < 0.01$.

and leisure effect (0.370); it had less influence on psychology and physiology (0.253) and water quality environment (−0.126). And found that environmental risk cognition and water quality environment showed a negative significance. The results verified that Hypotheses 5-1 and 5-2 were valid, but 5-3 was not.

It could be seen that environmental risk had a positive and significant influence on physical anxiety and leisure satisfaction, among which achievement performance and leisure effect had the greatest influence, and psychological and physiological influences were the least. Moreover, when the awareness of environmental risks increased, the public's satisfaction with the water quality and environment of the swimming pool decreased. As shown in **Table 3**.

Discussions

Analysis of Background

We believed that although swimming was a whole-body exercise, there were various swimming styles, and the exercise intensity of different swimming styles varied greatly. Swimming could strengthen cardiopulmonary function, effectively improve health (19, 20), help women maintain their posture, and provide people under 40 with sufficient exercise effects. In addition, most medical workers over the age of 51 work as supervisors or important units, and the medical work has heavy tasks. As a result, medical staff over 51 years old have less leisure time and less willingness to go out to swim. Therefore, under the epidemic environment, the majority of people who were still able to continue to participate in swimming were healthcare workers under the age of 40, and women were more willing to engage in swimming than men.

We believed that enterprises should strengthen the publicity and image of epidemic prevention measures and strengthen people's confidence in the epidemic prevention measures of swimming venues. Medical institutions could cooperate with enterprises nearby to provide medical personnel with a swimming and sports environment. Competent agencies could encourage medical personnel to plan swimming exercises to improve physical and mental health. All these measures would improve the willingness of men to participate and the willingness of medical staff of other ages to swim.

Analysis of Environmental Risk of Infection, Physical Anxiety, Leisure Satisfaction, and Physical and Mental Health of Swimmers

The research deduced that under the influence of the epidemic, public transportation and open environments were still at risk of infection (3). Therefore, people reduced their willingness to take public transportation and increased the travel time to the swimming pool. This led most people to believe that the existing exercise time was squeezed. However, the exercise time was insufficient, resulting in insufficient exercise, so the body was still tense, unable to solve the problems of insomnia, limb tension, and easy tantrums. In addition, due to China's current measures to clear and prevent blockages, the government used mobile phones to control personnel dynamics. Under the long-term epidemic prevention propaganda, the public's awareness of self-epidemic prevention and keeping distance gradually increased. In addition, the swimming pool used sodium hypochlorite for water quality management and environmental cleaning, and its items could effectively reduce the activity of the new coronavirus (25, 26), coupled with the repeated water filtration mechanism, to ensure the hygiene and quality of swimming pool water. Therefore, most healthcare workers believe that the swimming pool is in line with their anti-epidemic benefits and is a safe

and healthy leisure sports environment that can help relieve tension. The exercise investment could slow down, people could maintain their own pace to modify their swimming style and technique, and swimming could increase swimming ability and life performance.

In addition, in a low-risk environment, the mentality was stable. The exercise investment could slow down, people could maintain their own pace to modify their swimming style and technique, and swimming could increase swimming ability and life performance.

Therefore, most healthcare workers believed that swimming could modify body movements, but there was still a risk of infection. Although the performance of the exercise was satisfactory, the body was still tense, the muscles could not relax, and the exercise goal could not be achieved. Even if the distance of epidemic prevention and the clean water were reassuring, the exercise ability could be improved. However, issues such as wearing masks by swimmers, cleaning auxiliary facilities, or returning to the territory had not been perfected and could not be implemented. Ultimately, physical and mental health still produced feelings of lack of time, insomnia, and limb tension.

Next, different gender dimensions were explored. Due to the degree of exercise investment in different genders, there were certain differences in exercise time planning and immunity (75). Due to the differences in innate physiological structure, men had high exercise ability and high intensity, but their metabolism was slow, which led to muscle tension. In addition, the tasks of male healthcare workers will be heavier, increasing the pressure burden. As a result, it is difficult for them to recover from fatigue, sleep, and physical and mental stress.

Women were more delicate in their minds and were more able to find measures to disperse the stress and maintain their level of exercise when facing pressure. This is also the main reason why women are valued as healthcare workers. However, due to the diversity of swimming pool activity spaces and participants' identities, as well as different sanitation and epidemic prevention literacy, the risk of infection increased. When women engaged in sports in an uncertain sports environment, they easily lost the existing rhythm of life, disrupted existing plans, and affected life behavior and sports performance. As a result, the daily routine was changed, the way of eating was affected, the frustration and loss in life increased, and the physical and mental health were damaged.

Therefore, different genders had cognitive differences in physical anxiety physical and mental health cognition. Among them, men reported higher levels of tension, physical incoordination, head pressure, fatigue, and insomnia, while women experienced higher levels of issues such as transcendence, environmental stress, unsatisfactory performance, eating disorders, and worthlessness.

Finally, different age levels were discussed. Due to differences in reaction ability, virus resistance, life behavior, and mobility at different ages, different ages had different views on physical anxiety and physical and mental health cognition when faced with an infected sports environment (77, 78). Among them, because 31–40-year-old healthcare workers were in the process of physical maturity and strong strength, and the current work stage might be starting or reaching an important growth moment, there was an urgent need for ways to relieve stress. However, although swimming pools' pool management methods and environmental cleaning measures had a certain protective power, the epidemic situation continued to appear, and there was a gap between the public's personal epidemic prevention awareness and self-protection awareness, which was prone to epidemic prevention gaps. In addition, there was a lot of pressure from work or life, but it was impossible to relieve the pressure through sufficient exercise time and amount of exercise. As a result, people's physical and mental health in the 31–40-year-old age group had not been improved.

Furthermore, most people aged 41–50 had accumulated many years of experience in life, work, family care, and personal physical and mental health management (78). Most healthcare workers had a wealth of experience and knowledge at this age. Therefore, although there might be concerns that the current surrounding living environment was still full of the risk of virus infection, reducing the amount of exercise could not achieve the expected exercise goals. However, in the face of the epidemic environment and risks, people of this age group could plan their own health and epidemic prevention techniques and strategies based on their past experience and knowledge.

In addition, although people over the age of 51 were in the peak period of life, work, family, and self-health management (79), they might be about to stabilize or decline in another life stage (77, 80). However, at the same time, they were also gradually facing a state of physical exercise ability, self-health defense function, and mechanism gradually failing (78). Coupled with the impact of the epidemic, the convenience of transportation had disappeared, and the risk to the living environment had increased (3). As a result, people over the age of 51 could use their rich personal experience to find suitable alternative exercise methods and self-health management strategies under the epidemic situation to achieve the goal of stable life and exercise quality. However, due to the heavy medical work during the epidemic, coupled with the gradual failure of the individual's self-health defense function and mechanism, and the fear of virus infection, physical and mental health would be threatened, causing tension in the head, shoulders, neck, back, and other body parts and affecting sleep quality.

Therefore, different ages had different views on physical anxiety and physical and mental health cognition. Among them, 31–40-year-olds were more sensitive to issues such as satisfaction with physical anxiety, achievement of goals, surpassing levels, and physical and mental health, as well as fear, pain sensitivity, back tightness, stomach pain, eating disorders, and life burnout. People aged 41–50 were more sensitive to issues such as meeting their goals and facing environmental pressures on their physical and mental health. Most healthcare workers over the age of 51 were more sensitive to issues such as physical anxiety, fear of physical and mental health, head pressure, pain sensitivity, back tightness, and insomnia.

We suggested that a safe and proper transportation epidemic prevention mechanism should be planned, and appropriate transportation routes should be established. The study also recommended designing swimming and physical exercise

mechanisms in line with men, improving swimmers' epidemic prevention and hygiene literacy, and providing women with a safe swimming environment. Then, according to the needs of people aged 31–50, appropriate exercise prescriptions should be well planned, and exercise measures that could eliminate head and body fatigue, promote blood circulation, and improve the quality of sleep should be well designed to improve the health problems of people over 51 years old. The researchers believe that these measures can stabilize the health of healthcare workers and achieve the purpose of improving physical and mental health.

Correlation Analysis of Sports Environmental Risk on Physical Anxiety, Leisure Satisfaction, and Physical and Mental Health Cognition

This study inferred those studies had confirmed that the use of sodium hypochlorite for disinfection and cleaning in swimming pools could inhibit the activity of SARS-CoV-2 and achieve disinfection effects (27–30, 42), and swimming could help improve physical and mental health and strengthen cardiopulmonary function (19, 20). However, since the SARS-CoV-2 virus is still raging, Omicron has a low fatality rate and a high risk of infectivity. The emergence of mutant strains has the risk of re-infection, resulting in increased resistance to people trying to restore order in life and leisure. Coupled with the limited sports environment and people's different personal epidemic prevention sensitivities, hygiene, and environmental literacy, there might still be infection risks in swimming pools. This would lead to a reduction in the swimming people's leisure performance and final sports effect, and the expected leisure benefits and sports goals could not be obtained.

Therefore, the public's environmental risk cognition had a positive and significant correlation with physical anxiety and leisure satisfaction cognition. Among them, achievement performance and leisure effect were most influenced by environmental risk cognition. When the environmental risk cognition was higher, swimming pool water quality and environmental satisfaction would decrease.

We suggested that if the swimming pool environment could maintain the existing quality of disinfection and filtration, medical workers could first be provided with different leisure sports measures to achieve the purpose of improving physical and mental health. It would be a better decision if the swimming pool staff could repeatedly remind other swimmers to abide by the epidemic prevention regulations in the swimming pool, regulate the safe exercise distance, and continue to use sodium hypochlorite to increase disinfection in other swimming pool spaces.

CONCLUSIONS AND SUGGESTIONS

Medical workers are able to maintain the effect of exercise and maintain their health by strengthening the measures of water filtration and disinfection, enhancing the confidence in the swimming pool environment and epidemic prevention and control because the operators control the living footprints of swimmers. However, the public's awareness of epidemic prevention, hygiene awareness, and environmental literacy are different, causing people to have doubts about the effectiveness of epidemic prevention in bathrooms, toilets, and parking spaces. These lead effectively improve the environmental risk perception of medical personnel, reduce anxiety, improve leisure benefits, and achieve the goal of promoting physical and mental health. Female medical workers believe that the current epidemic prevention effect is still damaged, and male medical workers believe that the current sports benefits of going swimming are not good. Those aged 31–50 feel that exercise program options are few and ineffective, and those over 51 feel they still have problems with fear, head pressure, pain sensitivity, back tightness, and insomnia. The resulting environmental risk perceptions are positively correlated with body anxiety and leisure satisfaction perceptions and affect achievement performance and leisure outcomes. The higher the environmental risk, the more pronounced the impact of pool water quality and the effectiveness of environmental maintenance on healthcare workers.

We suggested that a safe and proper transportation epidemic prevention mechanism should be planned, and appropriate transportation routes should be established. Swimming pools can plan swimming exercise mechanisms for different genders and ages, strengthen publicity and epidemic prevention mechanisms, and improve the cooperation of other swimmers in epidemic prevention. These strategies will effectively improve the environmental risk perception of medical personnel, reduce anxiety, improve leisure benefits, and achieve the goal of promoting physical and mental health. Based on the above analysis and inference results, the study suggests research on the government, venues, the public, and future studies.

About the Government

Plan a safe and proper transportation epidemic prevention mechanism, and establish suitable transportation routes. Improve people's epidemic prevention and hygiene literacy, and provide a safe swimming environment.

About the Pool Venue

The swimming pool environment should maintain the existing disinfection and filtration quality, improve the epidemic prevention measures, add slogans or voice announcements to remind the public to abide by the epidemic prevention regulations in the swimming pool sports environment, and then standardize the safe exercise distance in the existing swimming sports space. Or, in the building space, add other sports facilities, and use the disinfection effect of sodium hypochlorite to maintain the cleanliness of the sports space and facilities.

About the Healthcare Workers

Improve environmental literacy, follow the epidemic prevention measures of the government and sports venues, jointly maintain the sports environment, maintain the effectiveness of epidemic prevention, and avoid becoming an epidemic prevention gap.

Suggestions for Future Research Directions

The researchers suggest extending the discussion to different occupations or identities of people, as well as other issues. It is expected that follow-up researchers will develop swimming exercise prescriptions or other physical exercise mechanisms in line with 31–50-year-old males and design exercise measures that can eliminate head and body fatigue, promote blood circulation, and improve sleep quality for people over 51 years old. The research gaps will be completed.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethics review and approval/written informed consent was not required as per local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

Z-CL and Y-HT: conceptualization and funding acquisition. Z-CL: methodology, software, visualization, resources, and project administration. Z-CL, C-HT, H-HL, and Y-HT: validation. C-HT: formal analysis. H-HL: investigation, data curation, writing—original draft preparation, and supervision. Y-HT: writing—review and editing. All authors have read and agreed to the published version of the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Tuncer FF. The spread of fear in the globalizing world: the case of COVID-19. *J Public Aff.* (2020) 20:e2162. doi: 10.1002/pa.2162
2. Das KP, Sharma D, Saha S, Satapathy BK. From outbreak of COVID-19 to launching of vaccination drive: invigorating single-use plastics, mitigation strategies, and way forward. *Environ Sci Pollut Res.* (2021) 28:55811–45. doi: 10.1007/s11356-021-16025-4
3. Cirrincione L, Plescia F, Ledda C, Rapisarda V, Martorana D, Moldovan RE, et al. COVID-19 epidemic: prevention and protection measures to be adopted at the workplace. *Sustainability.* (2020) 12:3603. doi: 10.3390/su12093603
4. Woods JA, Hutchinson NT, Powers SK, Roberts WQ, Gomez-Cabrera MC, Radak Z, et al. The COVID-19 epidemic and physical activity. *Sports Med Health Sci.* (2020) 2:55–64. doi: 10.1016/j.smhs.2020.05.006
5. Roviello V, Gilhen-Baker M, Vicidomini C, Roviello GN. Forest-bathing and physical activity as weapons against COVID-19: a review. *Environ Chem Lett.* (2021) 20:131–40. doi: 10.1007/s10311-021-01321-9
6. Sahu A, Naqvi WM. Quarantine exercises in the time of Covid-19-a review. *J Evolution Med Dent Sci.* (2020) 9:1922–7. doi: 10.14260/jemds/2020/418
7. Feng F, Tuchman S, Denninger JW, Fricchione GL, Yeung A. Qigong for the prevention, treatment, and rehabilitation of COVID-19 infection in older adults. *Am J Geriatr Psychiatry.* (2020) 28:812–9. doi: 10.1016/j.jagp.2020.05.012
8. Shen CC, Tseng YH, Shen MCS, Lin HH. Effects of sports massage on the physiological and mental health of college students participating in a 7-week intermittent exercises program. *Int J Environ Res Public Health.* (2020) 18:5013. doi: 10.3390/ijerph18095013
9. Huan PF, Chen CY, Lin HH, Wang IC, Hsu CH. In a stressful social environment, can using sports apps relieve the physical and mental stress of the elderly? From the perspective of Industry 4.0. *Math Probl Eng.* (2021) 1:1–14. doi: 10.1155/2021/4130972
10. Lin HH, Lin JW, Chen CC, Hsu CH, Lai BS, Lin TY. Effects of leisure obstacles, job satisfaction, and physical and mental health on job intentions of medical workers exposed to COVID-19 infection risk and workplace stress. *Healthcare.* (2021) 9:1569. doi: 10.3390/healthcare9111569
11. Chang KH, Hung CH, Lin TY, Lin HH, Chen YL. A study on the effects of Chinese massage on physical and mental health in participants based smart healthcare. *J Healthc Eng.* (2022) 1:1–13. doi: 10.1155/2022/9732355
12. Wang C, Zheng Y, Niu Z, Jiang X, Sun Q. The virological impacts of SARS-CoV-2 D614G mutation. *J Mol Cell Biol.* (2021) 13:712–20. doi: 10.1093/jmcb/mjab045
13. Zhou Y, Zhang L, Xie YH, Wu J. Advancements in detection of SARS-CoV-2 infection for confronting COVID-19 epidemics. *Lab Invest.* (2021) 102:4–13. doi: 10.1038/s41374-021-00663-w
14. World Health Organization. *Apart together Survey: Preliminary Overview of Refugees and Migrants Self-Reported Impact of COVID-19.* (2020). Available online at: https://apps.who.int/iris/bitstream/handle/10665/337931/9789240017924-eng.pdf (accessed January 31, 2022).
15. Olawale BE, Mutongoza BH, Adu EO, Omodan BI. COVID-19 induced psychosocial challenges in South African higher education: experiences of staff and students at two rural universities. *Res Soc Sci Technol.* (2021) 6:179–93. doi: 10.46303/ressat.2021.37
16. Centers for Disease Control, Ministry of Health and Welfare, Republic of China. *COVID-19.* Available online at: https://sites.google.com/cdc.gov.tw/2019ncov/global (accessed May 31, 2022).
17. Chen S. What implications does COVID-19 have on sustainable economic development in the medium and long terms? *Front Econ China.* (2020) 15:380–95. Available online at: https://ideas.repec.org/a/fec/journl/v15y2020i3p380-395.html
18. Sneppen K, Nielsen BF, Taylor RJ, Simonsen L. Overdispersion in COVID-19 increases the effectiveness of limiting nonrepetitive contacts for transmission control. *Proc Natl Acad Sci.* (2021) 118:e2016623118. doi: 10.1073/pnas.2016623118
19. Yang H, Ma J. How an epidemic outbreak impacts happiness: factors that worsen (vs. protect) emotional well-being during the coronavirus pandemic. *Psychiatry Res.* (2020) 289:113045. doi: 10.1016/j.psychres.2020.113045
20. Aşçi FH. The effects of physical fitness training on trait anxiety and physical self-concept of female university students. *Psychol Sport Exerc.* (2003) 4:255–64. doi: 10.1016/S1469-0292(02)00009-2
21. Kim KH, Lee BA, Oh DJ. Effects of aquatic exercise on health-related physical fitness, blood fat, and immune functions of children with disabilities. *J Exerc Rehabil.* (2018) 14:289–93. doi: 10.12965/jer.1836068.034
22. García-Ávila F, Valdiviezo-Gonzale L, Cadme-Galabay M, Gutiérrez-Ortega H, Altamirano-Cárdenas L, Zhindón-Arévalo C, et al. Considerations on water quality and the use of chlorine in times of SARS-CoV-2 (COVID-19) epidemic in the community. *Case Stud Chem Environ Eng.* (2020) 2:100049. doi: 10.1016/j.cscee.2020.100049
23. Hlavsa M, Beach MJ. *The Three Es of Healthy Swimming.* (2011). Available online at: https://www.cdc.gov/nceh/ehs/docs/jeh/2011/May_Hlavsa_Beach.pdf (accessed January 31, 2022).
24. Pickens SR. *Relative Effects of pH and Cyanurate on Disinfection.* (2019). Available online at: http://www.poolhelp.com/wp-content/uploads/2019/03/JSPSI-Volume-6-Number-1-pp-06-19-Pickins.pdf (accessed January 31, 2022).
25. Government of Alberta. *COVID-19. Guidance for Swimming Pools and Whirlpools.* (2021). Available online at: https://www.alberta.ca/assets/documents/covid-19-relaunch-guidance-swimming-pools-and-whirlpools.pdf (accessed January 31, 2022).

26. Siami Z, Aghajanian S, Mansouri S, Mokhames Z, Pakzad R, Kabir K, et al. Effect of ammonium chloride in addition to standard of care in outpatients and hospitalized COVID-19 patients: a randomized clinical trial. *Int J Infect Dis.* (2021) 108:306–8. doi: 10.1016/j.ijid.2021.04.043
27. Wang J, Shen J, Ye D, Yan X, Zhang Y, Yang W, et al. Disinfection technology of hospital wastes and wastewater: suggestions for disinfection strategy during coronavirus disease 2019 (COVID-19) epidemic in China. *Environ Pollut.* (2020) 262:114665. doi: 10.1016/j.envpol.2020.114665
28. World Health Organization. *Coronavirus Disease 2019 (COVID-19)—Section Water*. Available online at: https://www.cdc.gov/coronavirus/2019-ncov/faq.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fphp%2Fwater.html#COVID-19-and-Water (accessed January 31, 2022).
29. Pereira SSP, Oliveira HM de, Turrini RNT, Lacerda RA. Disinfection with sodium hypochlorite in hospital environmental surfaces in the reduction of contamination and infection prevention: a systematic review. *Rev Esc Enferm USP.* (2015) 49:0681–8. doi: 10.1590/S0080-623420150000400020
30. Köhler AT, Rodloff AC, Labahn M, Reinhardt M, Truyen U, Speck S. Efficacy of sodium hypochlorite against multidrug-resistant gram-negative bacteria. *J Hosp Infect.* (2018) 100:e40–6. doi: 10.1016/j.jhin.2018.07.017
31. Lin HH, Ling Y, Lin JC, Liang ZF. Research on the development of religious tourism and the sustainable development of rural environment and health. *Int J Environ Res Public Health.* (2021) 18:2731. doi: 10.3390/ijerph18052731
32. Nutsford D, Pearson AL, Kingham S. An ecological study investigating the association between access to urban green space and mental health. *Public Health.* (2013) 127:1005–11. doi: 10.1016/j.puhe.2013.08.016
33. Shen CC, Liang CF, Hsu CH, Chien JH, Lin HH. Research on the impact of tourism development on the sustainable development of reservoir headwater area using China's Tingxi reservoir as an example. *Water.* (2020) 12:3311. doi: 10.3390/w12123311
34. Wu CH. Research on the coastal marine environment and rural sustainable development strategy of island countries—taking the Penghu Islands as an example. *Water.* (2021) 13:1434. doi: 10.3390/w13101434
35. Clayton S, Karazsia BT. Development and validation of a measure of climate change anxiety. *J Environ Psychol.* (2020) 69:101434. doi: 10.1016/j.jenvp.2020.101434
36. Kurowski Ł. Household's overindebtedness during the COVID-19 crisis: the role of debt and financial literacy. *Risks.* (2021) 9:62. doi: 10.3390/risks9040062
37. Tseng KC, Lin HH, Lin JW, Chen IS, Hsu CH. Under the COVID-19 environment, will tourism decision making, environmental risks, and epidemic prevention attitudes affect the people's firm belief in participating in leisure tourism activities? *Int J Environ Res Public Health.* (2021) 18:7539. doi: 10.3390/ijerph18147539
38. Lin HH, Chang KH, Tseng CH, Lee YS, Hung CH. Can the development of religious and cultural tourism build a sustainable and friendly life and leisure environment for the elderly and promote physical and mental health? *Int J Environ Res Public Health.* (2021) 18:1989. doi: 10.3390/ijerph182211989
39. Giorgi G, Lecca LI, Alessio F, Finstad GL, Bondanini G, Lulli LG, et al. COVID-19-related mental health effects in the workplace: a narrative review. *Int J Environ Res Public Health.* (2020) 17:7857. doi: 10.3390/ijerph17217857
40. Hsu CH, Lin HH, Jhang S. Sustainable tourism development in protected areas of rivers and water sources: a case study of Jiuqu stream in China. *Sustainability.* (2020) 12:5262. doi: 10.3390/su12135262
41. Yu JH, Lin HH, Lo TC, Tseng KC, Hsu CH. Is the travel bubble under COVID-19 a feasible idea or not? *Int J Environ Res Public Health.* (2021) 18:5717. doi: 10.3390/ijerph18115717
42. World Health Organization. *Cleaning and Disinfection of Environmental Surfaces in the Context of Covid-19*. Available online at: https://apps.who.int/iris/bitstream/handle/10665/332096/WHO-2019-nCoV-Disinfection-2020.1-eng.pdf?sequence=1&isAllowed=y (accessed January 31, 2022).
43. Sánchez Y, Mehrvar M, McCarthy L, Quiñones E, Cheu L, Romero J. *Scientific Discussion of Selected Viruses Including COVID-19 in Hot Springs.* (2020). Available online at: https://ieeexplore.ieee.org/abstract/document/9240267/metrics#metrics (accessed January 31, 2022).
44. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 epidemic: a call for action for mental health science. *Lancet Psychiatry.* (2020) 7:547–60. doi: 10.1016/S2215-0366(20)30168-1
45. Zhao L, Hwang WY, Shih TK. Investigation of the physical learning environment of distance learning under COVID-19 and its influence on students' health and learning satisfaction. *Int J Distance Educ Technol.* (2021) 19:1–20. doi: 10.4018/IJDET.20210401.oa4
46. Jurecka A, Skucińska P, Gadek A. Impact of the SARS-CoV-2 coronavirus epidemic on physical activity, mental health and quality of life in professional athletes—a systematic review. *Int J Environ Res Public Health.* (2021) 18:9423. doi: 10.3390/ijerph18179423
47. Beck U. *Risk Society: Towards a New Modernity*. London: Sage Publications. (1992).
48. Adger WN, Safra de. Campos R, Mortreux C. Mobility, displacement and migration, and their interactions with vulnerability and adaptation to environmental risks In: McLeman R, Gemenne F, editors. *Routledge Handbook of Environmental Displacement and Migration*. London: Routledge (2018).
49. Weber EU. *Understanding Public Risk Perception and Responses to Changes in Perceived Risk*. Cambridge: Cambridge University Press. (2017).
50. Jun-Male Lai. *Make Your Financial Management More Secure*. Taipei: Taiwan Financial Research and Training Institute (2013).
51. Dryhurst S, Schneider CR, Kerr J, Freeman ALJ, Recchia G, Bles AM, et al. Risk perceptions of COVID-19 around the world. *J Risk Res.* (2020) 23:944–1006. doi: 10.1080/13669877.2020.1758193
52. Zhi-Xuan Zhang. *The Relationship Analysis of Environmental Risk Perception and Friendly Environmental Behavior: A Case Study of Seoul, Korea* (Master thesis). Taipei City University, Taipei (2019).
53. Roehr B. American psychiatric association explains DSM-5. *BMJ.* (2013) 346:f3591. doi: 10.1136/bmj.f3591
54. Zhang Y, Zhou B, Qiu J, Zhang L, Zou Z. Heart rate variability changes in patients with panic disorder. *J Affect Disord.* (2020) 267:297–306. doi: 10.1016/j.jad.2020.01.132
55. Guterstam A, Abdulkarim Z, Ehrsson HH. Illusory ownership of an invisible body reduces autonomic and subjective social anxiety responses. *Sci Rep.* (2015) 5:1–8. doi: 10.1038/srep09831
56. Davey GCL, Wells A. *Worry and it's Psychological Disorders: Theory, Assessment and Treatment*. New Jersey, NJ: John Wiley & Sons (2006).
57. Hillyard P, Tombs S. Beyond criminology? In: Davies P, Leighton P, Wyatt T, editors. *The Palgrave Handbook of Social Harm*. Palgrave MacMillan (2021). p. 11–36.
58. Awada M, Zhu R, Becerik-Gerber B, Lucas G, Southers E. An integrated emotional and physiological assessment for VR-based active shooter incident experiments. *Adv Eng Inform.* (2021) 47:101227. doi: 10.1016/j.aei.2020.101227
59. Williams RB, Slak-Valek N. Pokémon GO is serious leisure that increases the touristic engagement, physical activity and sense of happiness of players. *Inf Technol Tourism.* (2019) 21:515–33. doi: 10.1007/s40558-019-00153-2
60. Ellis GD, Jiang J, Freeman PA, Lacanienta A, Jamal T. Leisure as immediate conscious experience: foundations, evaluation, and extension of the theory of structured experiences. *J Leis Res.* (2020) 51:581–600. doi: 10.1080/00222216.2020.1754735
61. Li H, Lv L, Zuo J, Bartsch K, Wang L, Xia Q. Determinants of public satisfaction with an urban water environment treatment PPP project in Xuchang, China. *Sus Cities Soc.* (2020) 60:102244. doi: 10.1016/j.scs.2020.102244
62. Kim B, Kim S, King B. Religious tourism studies: evolution, progress, and future prospects. *Tourism Recreation Res.* (2020) 45:185–203. doi: 10.1080/02508281.2019.1664084
63. Jamal T, Budke C. Tourism in a world with epidemics: local-global responsibility and action. *J Tourism Futures.* (2020) 6:181–8. doi: 10.1108/JTF-02-2020-0014
64. Peterson EWF. *The Role of Population in Economic Growth*. New York, NY: Sage Open (2017). p. 1–15. doi: 10.1177/2158244017736094
65. Galli A, Iha K, Pires SM, Mancini MS, Alves A, Zokai G, et al. Assessing the ecological footprint and biocapacity of Portuguese cities: critical results for environmental awareness and local management. *Cities.* (2020) 96:102442. doi: 10.1016/j.cities.2019.102442
66. Ochieng PA. An analysis of the strengths and limitation of qualitative and quantitative research paradigms. *Probl Educ 21st Century.* (2009) 13:13–8. Available online at: http://oaji.net/articles/2014/457-1393665925.pdf
67. Berry N, Lobban F, Bucci S, A. qualitative exploration of service user views about using digital health interventions for self-management

in severe mental health problems. *BMC Psychiatry.* (2019) 19:35. doi: 10.1186/s12888-018-1979-1
68. Strauss & Corbin. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage (1990).
69. Janesick VJ. The choreography of qualitative research design: minuets,improvisations, and crystallization. In Denzin NK, Lincoln YS, editors. *Handbook of Qualitative Research*. Thousand Oak, CA: Sage (2000). p. 379–99.
70. Gursoy D, Jurowski C, Uysal M. Resident attitudes: a structural modeling approach. *Ann Tourism Res.* (2002) 20:79–105. doi: 10.1016/S0160-7383(01)00028-7
71. Lakshmanna K, Khare N. Constraint-based measures for DNA Sequence mining using group search optimization algorithm. *Int J Intell Eng Syst.* (2016) 9:91–100. doi: 10.22266/ijies2016.0930.09
72. Esentürk OK, Jurecka A, Skucińska P, Gadek A. Activity for their children with autism spectrum disorders during the novel coronavirus outbreak. *Int J Dev Disabil.* (2020) 67:446–57. doi: 10.1080/20473869.2020.1769333
73. Cypress B. Qualitative research methods: a phenomenological focus. *Dimens Crit Care Nurs.* (2018) 37:302–9. doi: 10.1097/DCC.0000000000000322
74. Ajzen I. The theory of planned behavior. *Organ Behav Human Decis Process.* (1991) 50:179–211. doi: 10.1016/0749-5978(91)90020-T
75. Bower I, Tucker R, Enticott PG. Impact of built environment design on emotion measured *via* neurophysiological correlates and subjective indicators: a systematic review. *J Environ psychol.* (2019) 66:101344. doi: 10.1016/j.jenvp.2019.101344
76. Li Z, Chen Q, Feng L, Rodewald L, Xia Y, Yu H, et al. Active case finding with case management: the key to tackling the COVID-19 epidemic. *Lancet.* (2020) 396:63–70. doi: 10.1016/S0140-6736(20)31278-2
77. Storesletten K, Telmer CI, Yaron. A consumption and risk sharing over the life cycle. *J Monet Econ.* (2004) 51:609–33. doi: 10.1016/j.jmoneco.2003.06.005
78. Edde M, Leroux G, Altena E, Chanraud S. Functional brain connectivity changes across the human life span: from fetal development to old age. *J Neurosci Res.* (2021) 99:236–62. doi: 10.1002/jnr.24669
79. Whiteley L, Brown L, Lally M, Heck N, van den Berg JJ. A mobile gaming intervention to increase adherence to antiretroviral treatment for youth living with HIV: development guided by the information, motivation, and behavioral skills model. *JMIR Mhealth Uhealth.* (2018) 6:e96. doi: 10.2196/mhealth.8155
80. Lewis S, Willis K, Bismark M, Smallwood N. A time for self-care? Frontline health workers' strategies for managing mental health during the COVID-19 pandemic. *SSM-Ment Health.* (2022) 2:100053. doi: 10.1016/j.ssmmh.2021.100053

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lu, Tseng, Lin, Perng and Tseng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 20 June 2022
doi: 10.3389/fpubh.2022.920956

Utility of Scoring System for Screening and Early Warning of Cervical Cancer Based on Big Data Analysis

Dan Hou[1†], *Binjie Yang*[1†], *Yangdan Li*[2] *and Ming Sun*[1*]

[1] *Department of Health Medicine, General Hospital of Northern Theater Command, Shenyang, China,* [2] *China Medical University, Shenyang, China*

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Roshan Fernandes, Nitte Mahalinga Adyanthaya Memorial Institute of Technology, India
Sravan Kumar Reddy, R G M College of Engineering and Technology, India

***Correspondence:**
Ming Sun
sunming@syjqzyy.org.cn

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *15 April 2022*
Accepted: *16 May 2022*
Published: *20 June 2022*

Citation:
Hou D, Yang B, Li Y and Sun M (2022) Utility of Scoring System for Screening and Early Warning of Cervical Cancer Based on Big Data Analysis. Front. Public Health 10:920956. doi: 10.3389/fpubh.2022.920956

Objective: To explore the utility of the scoring system for screening and early warning of cervical cancer based on big data analysis.

Methods: A total of 420 women undergoing physical examination in Shenyang from January 2021 to January 2022 were screened by convenient sampling as the study subjects. All females accepted the human papilloma virus (HPV) tests and thin-prep cytology test (TCT), a Rating Questionnaire for Screening and Early Warning of Cervical Cancer was developed, and a warning threshold was derived according to the scores of the questionnaire and the goodness of fit for the results of HPV+TCT tests. The patients were graded according to the threshold, and corresponding intervention strategies for patients of different grades were developed.

Results: Among the 420 people undergoing physical examination, 92 (21.90%) obtained scores $\geq$8 points, and 328 (78.10%) obtained scores $<$ 8 points; in diagnosing cervical cancer, the Rating Questionnaire for Screening and Early Warning of Cervical Cancer had an AUC value of 0.848, specificity of 97.22%, and sensitivity of 86.46%; after scientific intervention, HPV test results showed a significant decrease in both high-risk positive cases and low-risk positive cases ($p < 0.05$), and TCT results showed that there was a significant difference in the number of patients with CIN I before and after intervention ($p < 0.05$).

Conclusion: The scoring system for screening and early warning of cervical cancer based on big data analysis presents certain clinical value in the clinical screening of cervical cancer, which can further improve the screening coverage, is of great significance for the diagnosis and treatment of disease, and helps physician implement hierarchical diagnosis and treatment quickly and precisely.

Keywords: big data analysis, scoring system for early warning, cervical cancer screening, utility, intervention measures

INTRODUCTION

Cervical cancer is a malignant tumor disease that endangers women's health (1). Investigation and research (2) have shown that about 85% of cervical cancer patients are from developing countries, and there are approximately tens of thousands of new cervical cancer cases each year worldwide, seriously threatening women's health and life safety. According to years of basic experimental and epidemiological studies (3, 4), human papilloma virus (HPV) infection is an important factor causing the disease. Cervical cancer 3-step screening technique (step 1: cervical cytology and HPV test; step 2: colposcopy; step 3: histopathological examination) is currently the most widely used clinical diagnosis technique, and its clinical application provides an effective pathological evidence for the early diagnosis and treatment of cervical cancer in developed countries or regions (5). Early detection and treatment can effectively prevent the deterioration of the condition and minimize mortality. However, in countries and regions with relatively lagged economic and medical conditions, most cervical cancer patients are already in the advanced stage of the disease when diagnosed, leading to a lower cure rate and higher mortality. Therefore, searching for convenient and efficient means of cervical cancer screening can provide important auxiliary judgment indicators for clarifying and diagnosing the disease, and lay a strong scientific foundation for selecting clinical treatment options (6–8). There are deficiencies in every screening method in the existing medical context, and no single examination modality achieves the combination of high sensitivity with high specificity. Based on this, by reviewing a great amount of relevant materials of cervical cancer, a Rating Questionnaire for Screening and Early Warning of Cervical Cancer was developed in this trial to get the score of each patient and then develop corresponding intervention measures, thus improving the disease screening rate.

TABLE 1 | Clinical data of study subjects.

Clinical item	Proportion [%(*n*)]
Age	
18–26 years	13.10 (55/420)
27–40 years	28.33 (119/420)
41–60 years	42.62 (179/420)
>60 years	15.95 (67/420)
BMI value	
<18.5	10.95 (46/420)
18.5–24.9	40.24 (169/420)
25.0–29.9	34.29 (144/420)
≥30.0	14.52 (61/420)
Occupation	
Employee of organizations, enterprises, public institutions	10.00 (42/420)
Service personnel	32.38 (136/420)
Farmer	20.24 (85/420)
Worker	15.00 (63/420)
Self-employed entrepreneur	16.43 (69/420)
Others	5.95 (25/420)
Place of residence	
Rural area	57.86% (243/420)
Urban area	42.14% (177/420)
Educational degree	
Junior college and above	18.81% (79/420)
Senior high school	35.00% (147/420)
Junior high school and below	39.52% (166/420)
Illiteracy	6.67% (28/420)
Paying method of medical treatment	
Medical insurance	54.52 (229/420)
Self-paying	45.48 (191/420)
Family monthly income (yuan)	
≥3,000	49.52 (208/420)
<3,000	50.48 (212/420)

MATERIALS AND METHODS

General Data

A total of 420 women undergoing physical examination in Shenyang from January 2021 to January 2022 were screened by convenient sampling as the study subjects. The study met the World Medical Association Declaration of Helsinki (2013) (9) and reviewed and approved by the Hospital Ethics Committee.

Inclusion and Exclusion Criteria

Inclusion criteria: (1) Women with history of sexual life; (2) those with good understanding and could objectively and accurately fill in the questionnaire; and (3) those who voluntarily joined the study. Exclusion criteria: (1) Those who had cervical cancer and operation history; (2) those who were unable to fill in the questionnaire due to other reasons; and (3) those who were unwilling to join the study.

Methods

Content of Questionnaire

By reviewing a large number of published works on the epidemiology of cervical cancer and related influencing factors, the first version of Rating Questionnaire for Screening and Early Warning of Cervical Cancer was developed, which included the following sections. (1) The general data of the study subjects, such as their age, place of residence, occupation, educational degree, marital status, and the paying method of medical treatment;

(2) Sexual habits, such as age at first sexual activity, number of sexual partners, and whether condoms were used at the time of intercourse;

(3) Personal situation, such as number of times of miscarriages, number of times of pregnancies, presence or absence of an intrauterine contraceptive device (IUD), and presence or absence of immunodeficiency diseases; and

(4) Relevant symptoms or vital signs, such as presence or absence of cervical contact bleeding, cervical erosion, or cervical polyps.

In addition to the personal information of the study subjects, a total of 20 items were scored, and the maximum score was 20 points. The scoring method was as follows: 1 point was assigned to each question (1 for each forward-rating item, and −1 for each

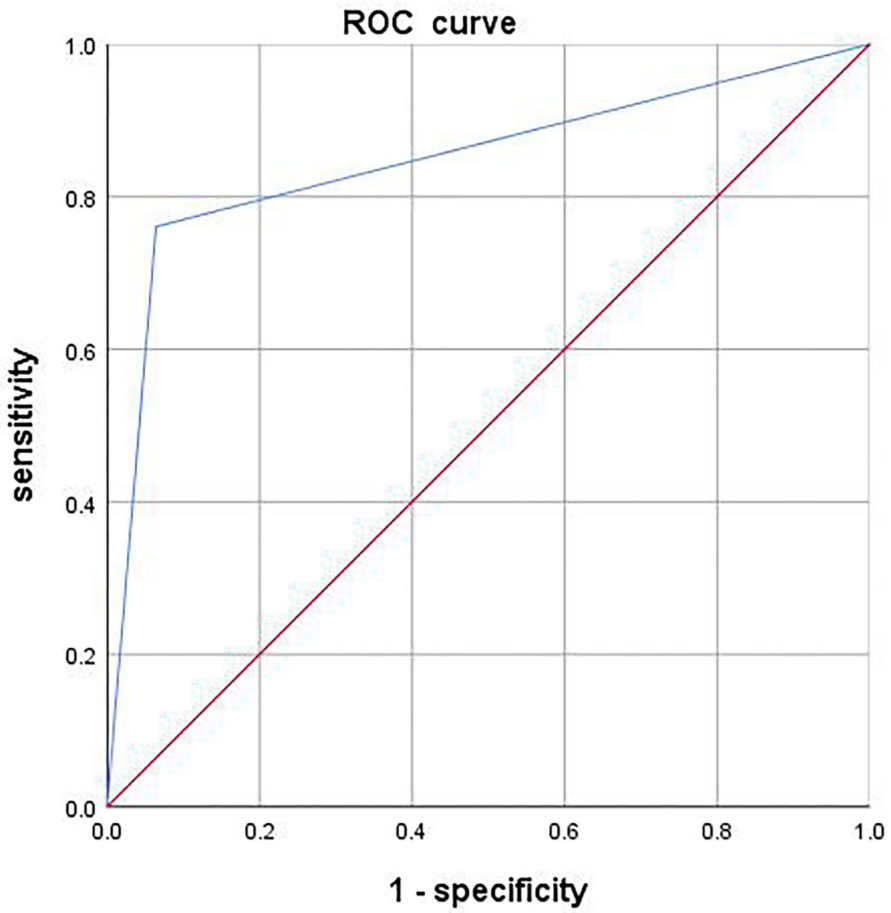


FIGURE 1 | Clinical diagnostic efficacy of the rating questionnaire for screening and early warning of cervical cancer.

backward-rating item), and higher scores indicated higher risk of illness of the study subjects.

Formation of Questionnaire

A Delphi Consultation Questionnaire was developed according to the corresponding general format. After 2 rounds of expert consultation and pre-survey to 30 females, relevant data were organized and the indicators were perfected in combination with the experts' suggestions, which then concluded that the Cronbach's α coefficient and split-half coefficient of the questionnaire were, respectively, 0.796 and 0.691, the validity index of each item level was ≥0.945, and the content validity index of the scale level was 0.950.

Test and TCTHPV Test

The HPV special sampler was inserted to the orificium externum isthmus of the subjects, rotated for 5 turns, and then slowly taken out and put into the HPV specimen bottle containing preserving liquid. The HPV subtyping genes chip was used to detect 26 types of HPV genes (such as HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 6, 11, 40, 42, 43, 44, 53, 54, 55, 57, 66, 67, 73, among which 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 were high-risk types, and 6, 11, 30, 42, 43, 44 were low-risk types), and all operations were conducted according to the specifications of the apparatus and reagent.

TCT

The sampler tip was inserted into the cervical canal and slowly rotated along the axis for 5 turns, the sampling brush tip was removed and put into the cell preservation solution, the specimen underwent standardized processing and was made into a thin-layer cell smear and fixed with 95% alcohol, and then Papanicolaou stain was performed. For diagnostic gradation of cytology, the Bethesda system (TBS) was adopted, to be specific, TCT negative, referring to no CLN was seen (negatives of intraepithelial lesions or malignancy, NILM); TCT abnormalities, such as a typical squamous cells of undetermined significance (ASC-US); low-grade squamous intraepithelial lesion (LSIL), which was graded as weak CIN (CIN I); high-grade squamous intraepithelial lesion (HSIL), which was graded as moderate and expressed CIN (CIN II and CIN III) and carcinoma *in situ* (CIS); and squamous cell carcinoma (SCC).

Quality Control

Each section of the investigation was under strict quality control, in the stage of designing, the scientificity and feasibility of the scheme and questionnaire applied were verified; before investigation, the researchers were trained with uniform guidelines, and then they provided guidance and correction on any queries and errors made by the study subjects when filling in the questionnaire; after the completion of the questionnaire, 2 personnel were assigned to collect and organize the questionnaires, Epidata3.1 was used for data entry, SPSS20.0 was used for statistical analysis, and an early warning threshold of 8 points for the questionnaire was derived by comparing the analytic findings and gynecological examination results.

Early Warning Threshold

According to patients' HPV test and TCT results and the scores of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer, the threshold was specified as 8 points, and detailed intervention measures were developed according to the threshold.

Intervention Strategies

Detailed hierarchical intervention measures were developed according to the threshold. (1) Below threshold: For those who were free from infection and without prior HPV testing, testing was recommended for 3 consecutive years, and if the results were negative, testing could then be done every 2–3 years; (2) Above threshold: this group was high-risk HPV infected individuals, who should visit gynecological clinics according to the TCT result, and the effectiveness of intervention measures should be analyzed by comparing the results of TCT and HPV test before and after the intervention.

Statistical Methods

The experimental data were statistically analyzed and processed by the software SPSS26.0, the diagnostic efficacy of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer on cervical cancer was analyzed by the area under the Receiver Operating Characteristic (ROC) curve.

RESULTS

Clinical Data of Study Subjects

See **Table 1**.

Scores of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer of the Females Undergoing Physical Examination

Among the 420 female undergoing physical examination, 92 (21.90%) obtained scores $\geq$ 8 points, and 328 (78.10%) obtained scores $<$ 8 points.

Clinical Diagnostic Efficacy of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer

See **Figure 1** and **Table 2**.

TABLE 2 | Comparison of the coincidence rate, specificity and sensitivity of early warning score diagnosis.

AUC value	S.E.	Specificity	Sensitivity	95% *CI* value
0.848	0.028	97.22%	86.46%	0.794-0.902

TABLE 3 | Comparison of various high-risk HPV infections in different cervical cancer lesions.

HPV type	CINI	CINII	CINIII	Cervical cancer
16	24	34	26	8
18	27	32	26	7
31	31	25	30	6
33	34	27	26	5
35	36	23	26	7
39	40	30	18	4
45	38	28	19	7
51	35	29	22	6
52	34	32	22	4
56	32	29	26	5
58	36	30	21	5
59	33	28	27	4
66	32	36	18	6
68	29	35	23	5

Comparison of Various High-Risk HPV Infections in Different Cervical Cancer Lesions

See **Table 3**.

Control Effect of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer in Cervical Cancer High-Risk Group

After scientific intervention, HPV test results showed a significant decrease in both high-risk positive cases and low-risk positive cases ($p < 0.05$), and TCT results showed that there was a significant difference in the number of patients with CIN I before and after intervention ($p < 0.05$), and no significant differences in other types were observed ($p > 0.05$). See **Table 4**.

DISCUSSION

Cervical cancer can seriously affect the quality of life of patients, lead to female infertility, and cause damage to the urinary system and vaginal bleeding, and in the late stage of the disease, cancer metastasis may occur, resulting in hydronephrosis as well as uremi. Data show (10, 11) that there are about 110,000 new cervical cancer cases in China each year. Screening of cervical cancer is able to detect precancerous lesions, and when the disease screening coverage rate is over 70%, the incidence and mortality risk of cervical cancer in the population can be effectively reduced. The pathogenesis of cervical cancer is HPV

TABLE 4 | Control effect on high-risk population above the threshold [n(%)].

Time	HPV		TCT			
	High-risk positive	Low-risk positive	CIN I	CINII	CINIII	Cervical cancer
Before intervention	35 (38.04%)	57 (61.96%)	28 (30.43%)	31 (33.70%)	24 (26.09%)	9 (9.78%)
After intervention	21 (22.83%)	36 (39.13%)	16 (17.39%)	22 (23.91%)	22 (23.91%)	8 (8.70%)
X^2	5.031	9.588	4.301	2.417	0.116	0.065
P	<0.05	<0.05	<0.05	0.143	0.733	0.799

infection, and multiparity and recurrent cervicitis is also the important causes of the high incidence of cervical cancer. The occurrence and development of cervical cancer is a gradual process, from years to decades, which has a long period of reversible precancerous lesions (12). Screening is an effective means to effectively reduce the incidence and mortality of cervical cancer, which can rapidly prevent invasive cervical cancer and precancerous changes in women, and therefore the World Health Organization (WHO) recommends that cervical cancer screening should be initiated worldwide for early diagnosis and early treatment (13). Developed countries have more mature technology and methods of cervical cancer screening and more complete screening programs, so their cervical cancer prevention and control is promising (14, 15). In 2014, HPV DNA typing assay was introduced to China as an initial screening method for cervical cancer, but HPV test cannot determine the specific disease type and single diagnosis has low sensitivity and poor specificity, also, TCT alone has a high rate of missed diagnoses. In addition, with China's large population base and limited health resources, the key point of cervical cancer screening is to develop specific screening programs and select appropriate screening methods (16, 17).

In this study, by reviewing a large number of published works on the epidemiology of cervical cancer and related influencing factors, the Rating Questionnaire for Screening and Early Warning of Cervical Cancer was developed and filled in by women undergoing physical examination, aiming to explore the utility of the questionnaire in diagnosing cervical cancer. Based on the imaging and scoring questionnaires of the study subjects, the specific threshold, i.e., the lowest or highest value an effect can produce, was developed, which helped physicians make a more precise judgment of the occurrence and progression of the disease. The content of the questionnaire included subjects' clinical data, high-risk factors of cervical cancer, and awareness. Previous studies have confirmed (18) that the risk of developing cervical cancer is 2.85 times higher in women having their first sexual intercourse at an age less than 20 years than those having their first sexual intercourse at an age over 20 years, so it is believed that having the first sexual intercourse at younger age is the main reason for the higher incidence of cervical cancer, as well as one of the important factors contributing to the occurrence of the disease in young women. In addition, the risk of cervical cancer in women with more than 3 times of childbirth is 5.91 times higher than that of those with less than 3 times of childbirth, demonstrating that fewer times of childbirth could lower the risk of cervical cancer to some extent (19). And studies also confirmed that low educational degree, poor economic level, poor sanitary conditions in rural areas, etc. are also the high-risk factors triggering cervical cancer. Women with lower education levels lack self-care awareness and gynecological knowledge, and because of their conservative traditional notion, they tend to conceal their illness and avoid treatment, thus increasing the risk of cervical cancer (20). The poor economic level and poor sanitation in rural areas lead to the lack of awareness of regular physical examination among women in these areas, which results in a high incidence of cervical cancer. Therefore, corresponding measures are required to pay close attention to people at high risk of cervical cancer, such as strengthening the publicity of female reproductive health knowledge, conducting targeted implementation of the popularization and publicity of health care knowledge to women, helping women in rural areas get good living habits, strengthening the basic medical and health conditions in township hospitals, putting more efforts in cervical cancer screening for women in rural areas, and assisting women in establishing a correct medical concept (21).

In this study, the diagnostic efficacy of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer in cervical cancer screening was analyzed by plotting ROC curves and taking the results of HPV test and TCT test as the "gold standard", and the results showed that the AUC value of the questionnaire was 0.848, which indicated a high-clinical application value. In addition, an early warning threshold for the questionnaire was derived in this study by analyzing the questionnaire scores and the gynecological examination findings of the females undergoing physical examination, thereby developing detailed hierarchical interventions based on the threshold. The study implemented scientific intervention measures to those above the threshold, and analyzed and compared the experimental results before and after intervention to explore the control effect. HPV typing assay is designed to target the DNA (the genetic material), through a process of PCR target preference, target amplification, automated real-time detection, and selective amplification, and then identify the corresponding HPV replication with greater precision than second-generation hybrid capture assay. Currently, TCT is the more advanced and mature screening technique for cervical cancer, but it has certain false-negative rate and false-positive rate for low-grade cervical lesions, and the concordance rate increases with higher cytological grade. Scientific means of clinical screening can screen out people at high risk of

cervical cancer, reduce the follow-up interval, facilitate effective monitoring of cervical cytological alterations and early detection of cervical lesions, and TCT combined with HPV screening is beneficial to the triage management of cervical lesions, realizing early detection and treatment of the disease. In this study, more detailed prevention and treatment strategies were developed by typing the obtained examination findings. The findings showed that after intervention, both high-risk positive cases and low-risk positive cases were significantly reduced according to the HPV test results ($p < 0.05$), and TCT results denoted that the number of patients with CINI was greatly reduced, which illustrated that the implementation of scientific control measures could lower the risk of cervical cancer to a certain extent. Studies have confirmed (22) that different degrees of TCT and different types of HPV also have significant differences in disease reversal cycles. Data from a large sample survey showed that patients with a TCT result of CIN I and low-risk positive HPV infection had a readily reversible condition, whereas those with a TCT result of CIN II and CIN III and high-risk positive HPV infection required further colposcopy examination and thus had a longer cycle of clinical intervention. Some scholars believe (23) that among multiple high-risk types of HPV, types 16 and 18 have a higher carcinogenic risk, which may be related to the strong ability of gene binding to the host cell, and their gene expression products can also affect the host cell cycle and apoptosis, leading to carcinogenesis. However, other types of viruses are prone to cause low-grade intraepithelial lesions, which may be related to the poor tolerance of the virus to the acidic environment of the vagina, poor viability, and weak pathogenic ability. Type 52 virus has a long cycle of negative conversion and obtains less effective result in clinical treatment; it is also the main genotype responsible for the development of cervicitis. The E6 and E7, biomarkers of cervical cancer caused by type 52 virus, often lead to base mutations that may exert an important function in the progression of cervical intraepithelial lesion and invasive carcinoma, and they are considered as the important carcinogens because they can inactivate the host tumor suppressor p53 and Rb (24).

In addition, unlike the general prevention and control of cervical cancer in developed countries, the comprehensive task of controlling cervical cancer in China is tough and difficult in implementation, and requires long time because of the large population and the uneven development of medical and health resources. Therefore, based on the current situation, the prevention and control measures in China can be summarized as the following 3 aspects. (1) Strengthen the infrastructure construction of the health system, invest corresponding financial, material, and human resources, strengthen the training efforts of primary care workers, and establish a complete system for AI-assisted diagnosis; (2) based on the domestic conditions, clarify the age and screening interval of subjects, identify high-risk individuals, encourage those of proper ages to actively participate in screening, establish a scientific surveillance and evaluation system, and encourage the population of the right age to get an HPV vaccination to gradually expand vaccination coverage; and (3) scientifically develop appropriate screening programs and strategies for the Chinese people to reduce the incidence of cervical cancer (25). Despite the progress of modern medical diagnosis technology, there are still many dilemmas in cervical cancer screening. As medical workers, we still need to continuously try our best to explore more efficient clinical technology, optimize the current diagnosis mode, and provide reliable data support for early screening of cervical cancer. In addition, the application of the Rating Questionnaire for Screening and Early Warning of Cervical Cancer greatly conserves medical and health resources, which is more common and low-cost. The scoring content of the questionnaire should be continually optimized in the future, so that it can be universal in various regions of China as a primary screening method for cervical cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of General Hospital of Northern Theater. PLA. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DH, BY, and MS: conception and design and data analysis and interpretation. All authors: administrative support, provision of study materials or patients, collection and assembly of data, manuscript writing, and final approval of manuscript.

FUNDING

This research was funded by Feasibility study on cervical cancer screening early warning scoring system for female physical examination population in Shenyang, Grant No. 20-205-4-088.

REFERENCES

1. Dudea-Simon M, Dudea S, Ciortea R, Malutan A, Mihu D. Elastography of the uterine cervix in gynecology: normal appearance, cervical intraepithelial neoplasia and cancer. A systematic review. *Med Ultrason.* (2021) 23:74–82. doi: 10.11152/mu-2646
2. Shou H, Yasuo Y, Yuan S, Lou H, Ni J. Association of pretreatment SUV of cervix and SCC-antigen with FIGO2018 stage in Stage IIB-IVB squamous cervical cancer and relationship to prognosis. *Int J Gynaecol Obstet.* (2021) 152:112–7. doi: 10.1002/ijgo.13465
3. Colacurci N, Schettino MT, Grimaldi V, De Luca FP, Mansueto G, Costa D, et al. Flow cytometry characterization of pluripotent transmembrane

glycoproteins on resident cervix uteri cells in patients screened for cervical cancer. *Cancer Invest.* (2020) 38:228–39. doi: 10.1080/07357907.2020.1742349
4. Dostalek L, Runnebaum I, Raspagliesi F, Vergote I, Dusek L, Jarkovsky J, et al. Oncologic outcome after completing or abandoning (radical) hysterectomy in patients with cervical cancer and intraoperative detection of lymph node positivity; ABRAX (ABandoning RAd hyst in cerviX cancer). *Int J Gynecol Cancer.* (2020) 30:261–4. doi: 10.1136/ijgc-2019-000890
5. Cea García J, de la Riva Pérez PA, Rodríguez Jiménez I, Márquez Maraver F, Polo Velasco A, Jiménez Gallard J, et al. Selective biopsy of the sentinel node in cancer of cervix: experience in validation phase. *Rev Esp Med Nucl Imagen Mol.* (2018) 37:359–65. doi: 10.1016/j.remnie.2018.07.002
6. Hashiguchi M, Kai K, Nishiyama S, Nakao Y, Yokoyama M, Aishima S. Clear cell carcinoma of the uterine cervix presented as a submucosal tumor arising in a background of cervical endometriosis. *Int J Gynecol Pathol.* (2018) 37:88–92. doi: 10.1097/PGP.0000000000000386
7. Chaman A, Inayat K, Hanif A, Afsar S, Ullah I. A study of agreement between visual inspection with acetic acid of cervix and pap smear for cervical cancer screening. *J Ayub Med Coll Abbottabad.* (2018) 30:S668–S670.
8. Wei B, Mei P, Huang S, Yu X, Zhi T, Wang G, et al. Evaluation of the SureX HPV genotyping test for the detection of high-risk HPV in cervical cancer screening. *Virol J.* (2020) 17:171. doi: 10.1186/s12985-020-01417-8
9. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. (2013) 310:2191–4. doi: 10.1001/jama.2013.281053
10. Zorzi M, Del Mistro A, Giorgi Rossi P, Laurino L, Battagello J, Lorio M, et al. Risk of CIN2 or more severe lesions after negative HPV-mRNA E6/E7 overexpression assay and after negative HPV-DNA test: Concurrent cohorts with a 5-year follow-up. *Int J Cancer.* (2020) 146:3114–23. doi: 10.1002/ijc.32695
11. Gottschlich A, van Niekerk D, Smith LW, Gondara L, Melnikow J, Cook DA, et al. Assessing 10-Year Safety of a Single Negative HPV Test for Cervical Cancer Screening: Evidence from FOCAL-DECADE Cohort. *Cancer Epidemiol Biomarkers Prev.* (2021) 30:22–9. doi: 10.1158/1055-9965.EPI-20-1177
12. Kang M, Ha SY, Cho HY, Chung DH, Kim NR, An J, et al. Comparison of papanicolaou smear and human papillomavirus (HPV) test as cervical screening tools: can we rely on HPV test alone as a screening method? An 11-year retrospective experience at a single institution. *J Pathol Transl Med.* (2020) 54:112–8. doi: 10.4132/jptm.2019.11.29
13. Banik R, Naher S, Rahman M, Gozal D. Investigating bangladeshi rural women's awareness and knowledge of cervical cancer and attitude towards HPV vaccination: a community-based cross-sectional analysis. *J Cancer Educ.* (2022) 37:449–60. doi: 10.1007/s13187-020-01835-w
14. Onuma T, Kurokawa T, Shinagawa A, Chino Y, Yoshida Y. Evaluation of the concordance in HPV type between self- and physician-collected samples using a brush-based device and a PCR-based HPV DNA test in Japanese referred patients with abnormal cytology or HPV infection. *Int J Clin Oncol.* (2020) 25:1854–60. doi: 10.1007/s10147-020-01727-5
15. Derbie A, Mekonnen D, Woldeamanuel Y, Van Ostade X, Abebe T, HPV. E6/E7 mRNA test for the detection of high grade cervical intraepithelial neoplasia (CIN2+): a systematic review. *Infect Agent Cancer.* (2020) 15:9. doi: 10.1186/s13027-020-0278-x
16. Garcés-Palacio IC, Sanchez GI, Baena Zapata A, Sánchez VC, Cosme Urrea Y, Zabala Rodríguez D et al. Psychosocial impact of inclusion of HPV test on the management of women with atypical squamous cells of undetermined significance: a study within a randomised pragmatic trial in a middle-income country. *Psychol Health.* (2020) 35:750–69. doi: 10.1080/08870446.2019.1678749
17. Locklar LRB, Do DP. Rural-urban differences in HPV testing for cervical cancer screening. *J Rural Health.* (2022) 38:409–15. doi: 10.1111/jrh.12615
18. Orumaa M, Leinonen MK, Campbell S, Møller B, Myklebust TÅ, Nygård M. Recent increase in incidence of cervical precancerous lesions in Norway: Nationwide study from 1992 to 2016. *Int J Cancer.* (2019) 145:2629–2638. doi: 10.1002/ijc.32195
19. Gnade CM, Hill EK, Botkin HE, Hefel AR, Hansen HE, Sheets KA, et al. Is the age of cervical cancer diagnosis changing over time? *J Gynecol Obstet Hum Reprod.* (2021) 50:102040. doi: 10.1016/j.jogoh.2020.102040
20. Li T, Yoon J, Luck J, Bui LN, Harvey M. The impact of continuity of care on cervical cancer screening: how visit pattern affects guideline concordance. *J Cancer Educ.* (2022) 37:30–6. doi: 10.1007/s13187-020-01777-3
21. Wang YY, Wang ZJ, Zhang Y, Gao XH, Yang CX, Zhao FH, et al. Economic evaluation of fifteen cervical cancer screening strategies in rural China. *Zhonghua Fu Chan Ke Za Zhi.* (2019) 54:840–7. doi: 10.3760/cma.j.issn.0529-567x.2019.12.008
22. Veena MS, Raychaudhuri S, Basak SK, Venkatesan N, Kumar P, Biswas R, et al. Dysregulation of hsa-miR-34a and hsa-miR-449a leads to overexpression of PACS-1 and loss of DNA damage response (DDR) in cervical cancer. *J Biol Chem.* (2020) 295:17169–86. doi: 10.1074/jbc.RA120.014048
23. Kadian LK, Yadav R, Nanda S, Gulshan G, Sharma S, Yadav C. High-risk HPV infection modulates the promoter hypermethylation of APC, SFRP1, and PTEN in cervical cancer patients of North India. *Mol Biol Rep.* (2020) 47:9725–32. doi: 10.1007/s11033-020-05960-z
24. Alam Z, Dean JA, Janda M. What do South Asian immigrant women know about HPV, cervical cancer and its early detection: a cross-sectional Australian study. *J Migr Health.* (2022) 5:100102. doi: 10.1016/j.jmh.2022.100102
25. Suchonska B, Gajewska M, Madej A, Wielgoś M. Cervical intraepithelial neoplasia during pregnancy. *Indian J Cancer.* (2020) 57:31–5. doi: 10.4103/ijc.IJC_403_18

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Hou, Yang, Li and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 20 June 2022
doi: 10.3389/fpubh.2022.894920

VCNet: Hybrid Deep Learning Model for Detection and Classification of Lung Carcinoma Using Chest Radiographs

Ritu Tandon[1]*, *Shweta Agrawal*[1]*, *Arthur Chang*[2]* and *Shahab S. Band*[3]*

[1] Institute of Advance Computing, SAGE University, Indore, India, [2] Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology, Douliu, Taiwan, [3] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliu, Taiwan

OPEN ACCESS

Edited by:
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

Reviewed by:
Ashutosh Kashiv,
Shivajirao Kadam Institute of Technology and Management
Indore, India
A. H. Alamoodi,
Sultan Idris University of Education, Malaysia
Joseph Henry Anajemba,
Hohai University, China
Muhammad Tariq Sadiq,
University of Lahore, Pakistan

***Correspondence:**
Ritu Tandon
bansalritu10@gmail.com
Shweta Agrawal
shweta_agrawal1883@rediffmail.com
Arthur Chang
changart@yuntech.edu.tw
Shahab S. Band
shamshirbands@yuntech.edu.tw

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 12 March 2022
Accepted: 22 April 2022
Published: 20 June 2022

Citation:
Tandon R, Agrawal S, Chang A and Band SS (2022) VCNet: Hybrid Deep Learning Model for Detection and Classification of Lung Carcinoma Using Chest Radiographs. Front. Public Health 10:894920. doi: 10.3389/fpubh.2022.894920

Detection of malignant lung nodules from Computed Tomography (CT) images is a significant task for radiologists. But, it is time-consuming in nature. Despite numerous breakthroughs in studies on the application of deep learning models for the identification of lung cancer, researchers and doctors still face challenges when trying to deploy the model in clinical settings to achieve improved accuracy and sensitivity on huge datasets. In most situations, deep convolutional neural networks are used for detecting the region of the main nodule of the lung exclusive of considering the neighboring tissues of the nodule. Although the accuracy achieved through CNN is good enough but this models performance degrades when there are variations in image characteristics like: rotation, tiling, and other abnormal image orientations. CNN does not store relative spatial relationships among features in scanned images. As CT scans have high spatial resolution and are sensitive to misalignments during the scanning process, there is a requirement of a technique which helps in considering spatial information of image features also. In this paper, a hybrid model named VCNet is proposed by combining the features of VGG-16 and capsule network (CapsNet). VGG-16 model is used for object recognition and classification. CapsNet is used to address the shortcomings of convolutional neural networks for image rotation, tiling, and other abnormal image orientations. The performance of VCNeT is verified on the Lung Image Database Consortium (LIDC) image collection dataset. It achieves higher testing accuracy of 99.49% which is significantly better than MobileNet, Xception, and VGG-16 that has achieved an accuracy of 98, 97.97, and 96.95%, respectively. Therefore, the proposed hybrid VCNet framework can be used for the clinical purpose for nodule detection in lung carcinoma detection.

Keywords: capsule network, convolutional neural networks, CT, MobileNet, VCNet, VGG-16, Xception

INTRODUCTION

Cancer is still regarded as a dangerous disease with severe death rates. Lung cancer has the highest mortality or death rate of any cancer and is widely regarded as the deadliest carcinoma among all types of cancer. Consequently, several researchers are focusing on ways of detecting lung cancer nodules through digital images, specifically through computed tomography (CT). CT scans use

X-rays to generate several images and create a challenge for radiologists to detect tiny nodules from these images (1). Analysis of nodules and their interpretation is the basic task performed by the radiologist for the diagnosis of lung cancer. Many scientists and researchers are working on automated solutions that will help doctors save time and money (2).

In most cases, the size and appearance of nodules give the first inference regarding cancer and can be classified as benign or malignant. Generally, lung nodules <3 cm are considered normal nodules, and those larger than 3 cm are considered malignant or lung masses. **Figures 1A,B** shows the images of benign and malignant lung cancer. Based on nodule classification and other findings, cancer probability can be assessed. AI techniques are playing a very important role in the primary level detection and classification of various types of cancer (3). A variety of disciplines, including medicine, agriculture, games, and many more, have benefited from the deployment of deep learning (DL) models in recent years. In all of these fields, DL models perform quite well, particularly in specific types of applications such as classification of images, identification of objects, and image segmentation (4). DL is a subfield of artificial intelligence, having interconnected nodes to perform complicated tasks. Instead of using pre-programmed instructions, DL algorithms are capable of learning from the training data. Many of the researchers have already worked for lung carcinoma detection using deep learning (5, 6).

In computer vision and radiography, convolutional neural networks (CNN) are a class of artificial neural networks that are gaining a lot of popularity. CNN is used to automatically detect features from the images using a number of layers like convolution layers, pooling layers, and fully connected layers (7, 8). CNN uses successive convolution and pooling layers to classify the images. The pooling layer in the CNN reduces the dimension and classifies the object regardless of its spatial information. That means where the object is actually located in the image (9). This pooling function of CNN is both an advantage and a drawback. During the pooling function, it loses some important information that is very useful in image segmentation and object detection. To overcome this drawback, Geoffrey Hinton (10) proposed capsule network architecture.

A capsule is a cluster or group of neurons that stores information about a specific item in a picture; the information is mainly about its position, rotation, scale, etc. in a high-dimensional vector (8 dimensions or 16 dimensions), each. The capsule network architecture is divided into three parts. (1) primary capsule (convolution, reshape, and squash functions); (2) higher level capsule (dynamic routing); and (3) loss functions (margin and reconstruction loss).Capsule networks (CapsNet) (10) are basically used to overcome the loss of information that comes from the pooling operation of CNN and obtain spatial information. CapsNet's multiple convolutional layers are wrapped in capsules.

This paper utilizes transfer learning and presents a hybrid model, VCNet, for detecting the lung nodules from CT scans. Binary classification is done using the input of lung CT scans and the output is either "benign" or "malignant". A lung CT image dataset is collected from LIDC-IDRI (11) and a new hybrid model that is a combination of VGG-16 and Capsule network is developed and applied for the classification and identification of lung cancer.VGG16 is one of the finest vision model architectures based on CNN computer vision techniques. The accuracy achieved by this model for ImageNet test data is 92.7%. The model presents an improved form of AlexNet and the main architectural changes are in filter size. The 16 in VGG is used to refer that it has 16 weight layers (12).

Related Work

Lung cancer with the greatest death rate is the most prevalent and aggressive form in India. Patients with non-small cell lung cancer (NSCLC) have a 5-year survival rate of 18%. According to the 2018 GLOBOCON report (1), 11.6 percent of all lung cancer cases and 18.4 percent of all lung cancer deaths. Patients with lung cancer outnumber those with other cancers such as breast, liver, cervical, skin, and so on. Patients with pulmonary cancer can enhance their survival chances if they discover early lung cancer. Detection of cancer at last stage does not permit surgical treatment in most of the cases, so last stage lesions are treated by nonsurgical treatment like chemotherapy, Radiotherapy and immunotherapy. Computer vision methods proved an important result in lung nodule classification and detection at an early stage through CT scans. For converting the raw input into the desired features end to end learning allows direct mapping with elimination of hand –crafted features. Machine learning algorithms provide extremely good outcomes in computer vision technology and analysis of medical images. Two main classes, massive training artificial neural networks (MTANNs) and revolutionary neural networks may be performed with end-to-end training (CNNs).Nima Tajbakhsh et al. utilized both theoretical and practical approaches to compare these two classes of end-to-end learning for the detection of lung nodules and the differentiation of benign and malignant tumors in low-dose CT scans. For the analysis of the same the authors have used four MTANN architectures and two CNN architectures that all are having different depth. They used a substantial dataset for CNN training in the second phase, and their results demonstrate a smaller performance gap between MTANN and CNN. After both theoretical and experimental approach they had presented that MTANN gives better performance as compared to the CNN (13).

Goran Jakimovski et al. presented double CNN and regular CNN with a max pooling layer for identifying cancer stages through CT scans (14). One of the most difficult tasks in radiology imaging is automatically determining the precise position of the lung nodule. Hongtao Xie et al. presented the automatic detection of lung nodules using 2D CNN for the reading process of CT scans. The authors did their experiments on the LUNA-16 dataset and achieved a sensitivity of 86.42% (15). By combining information from PET and CT scans, Qin et al. (16) demonstrated a DL architecture for the noninvasive detection of lung cancer. The authors achieved a 0.92 area under the ROC curve. Nakrani et al. (17) presented a method for nodule detection using Resnet architecture in CT scans. The dataset used here is LIDC and the accuracy achieved is 95.24% (17).

Lu et al. (18) proposed an optimal methodology for detection of lung carcinoma at early stage. For good network accuracy & optimal arrangement Marine predators' algorithm is used.

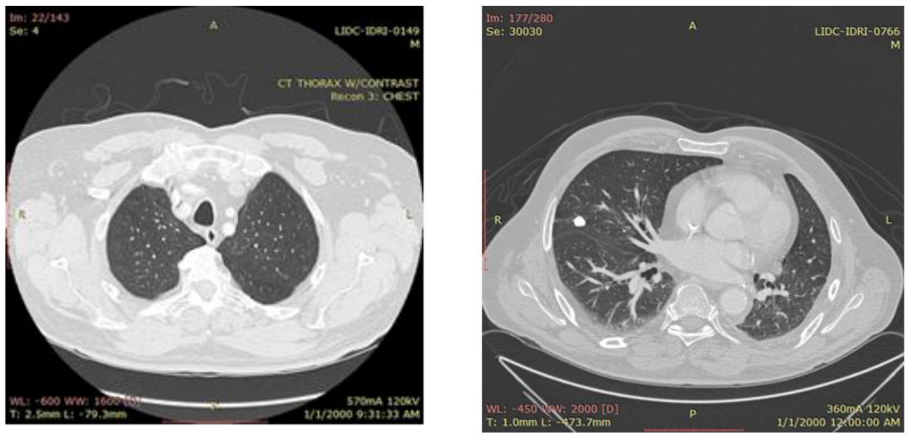


FIGURE 1 | (A) Benign lung image. **(B)** Malignant lung.

TABLE 1 | Summary of the research articles on lung cancer diagnosis and classification using the DL model by CNN.

Reference	Year	Purpose of study	Image acquisition	Image dataset	Accuracy/ sensitivity	Major findings/limitations
Jakimovski et al. (14)	2019	Lung cancer stage detection	CT images	LONI database	99.63%	A deep neural network has been designed to detect the stage of lung cancer. The pre-classification of photos is done using the K-mean approach
Xie et al. (15)	2019	Automated nodule detection	CT images	LUNA-16	86.42%	An automated pulmonary nodule detection-based model is proposed using two-dimensional CNN
Qi et al. (16)	2020	Classification of lung cancer	PET and CT images	Department of Radiology of the Henan Provincial People's Hospital	72% (Balanced Accuracy)	Used CT and PET multimodality noninvasive clinical images for lung cancer classification
Nakrani et al. (17)	2020	Lung nodule detection	CT images	LIDC-IDRI	95.24%	ResNet architecture of 2D convolutional neural network is used for detection of lung nodule
Neal Joshua et al. (19)	2021	Lung nodule detection	CT images	LUNA-16 dataset	97.17%	The 3D AlexNet architecture is being used. The 3D multi-dimensional convolutional neural network is used for lung nodule detection
Bharati et al. (20)	2020	Lung disease detection	X-ray images	Chest X-ray dataset of NIH collected from kaggle	73%	Proposed a hybrid model for lung disease detection using a modified capsule network
Afshar et al. (21)	2020	Lung nodule malignancy prediction	CT images	LIDC-IDRI	83%	The 3D multi-dimensional convolutional neural network is used for lung nodule detection

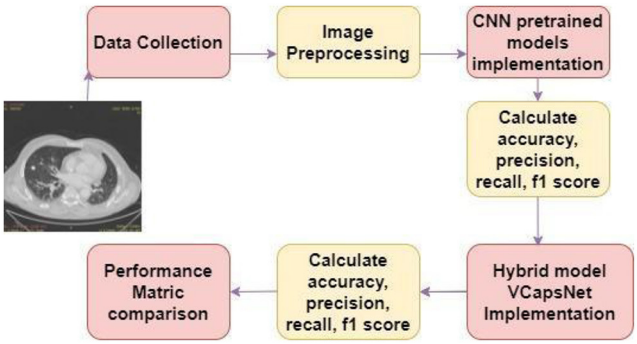


FIGURE 2 | Methodology.

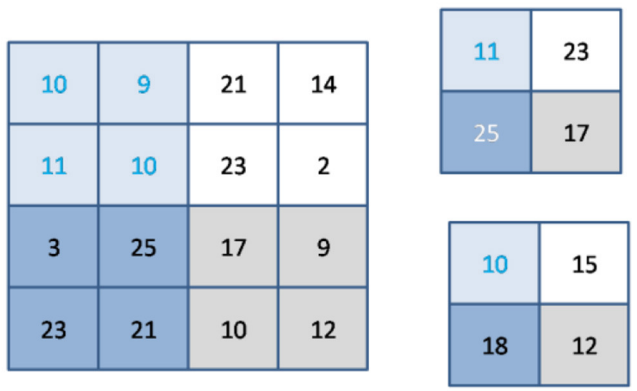


FIGURE 3 | Example of max & average pooling.

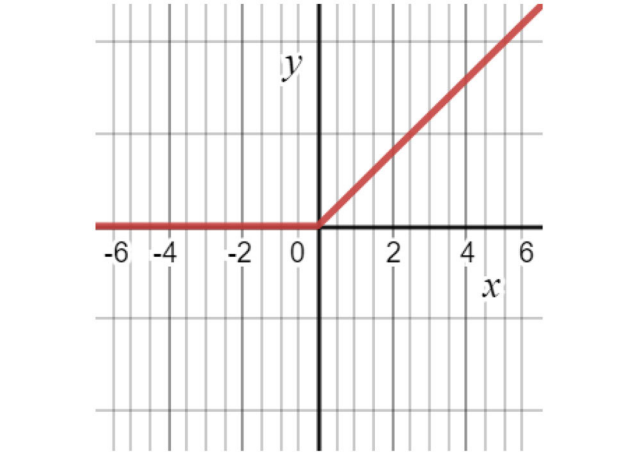


FIGURE 4 | ReLu activation function.

Table 1 shows a summary of the research articles on lung cancer diagnosis and classification using the DL model CNN.

Problem Statement

Many researchers have already presented various DL models for lung cancer detection, but the required accuracy on public datasets is still a big challenge for using those models for clinical purposes and also for radiologists to diagnose the cancerous nodules. Some of the researchers have achieved good accuracy in their work, but they have used either private datasets collected from hospitals or some small datasets available on the internet. In Tajbakhsh et al. (13), 95% accuracy was attained by comparing two end-to-end classes that are large artificial neural network training and convolutional neural networks. But for the implementation of the same, they used the unpublished dataset of 31 patients, including 38 scans collected from the participation of a lung cancer screening program. In (19), the LUNA-16 dataset was used to detect the lung nodules using the 3D AlexNet architecture of CNN and achieved an accuracy of 97.17%. In Jakimovski and Davcev (14), the achieved accuracy was 99.63% for lung cancer stage detection, but the authors of this paper used the private dataset of the University of South California.

To the best of our knowledge, the paper (20) implemented a hybrid model using a modified capsule network for lung disease detection and achieved an accuracy of 73%. For the detection of lung nodules, a multi-scale capsule network is proposed in (21) and achieved an accuracy of 83% for the LIDC dataset.

Based on these works the major contributions of our work are:

- A hybrid model named as VCNet is proposed by combining the features of VGG-16 and capsule network (CapsNet).
- A pre-trained VGG-16 model is used for feature extraction.
- CapsNet layers are used to handle misalignment problems with the existing deep learning models.
- Fully connected layers along with dropouts and sigmoid activation functions are also used to prevent overfitting and to build a generalized model.
- The performance of VCNet is verified on the Lung Image Database Consortium (LIDC) image collection dataset.

TABLE 2 | Model architecture.

S.No.	Model	Size	TOP 1 accuracy	TOP 5 accuracy	No of parameters
1	MobileNet	17 MB	0.665	0.871	4,253,864
2	Xception	88 MB	0.790	0.945	22,910,480
3	VGG-16	528 MB	0.715	0.901	138,357,544
4	ResNet	99 MB	0.759	0.929	25,636,712
5	Inception V3	92 MB	0.788	0.944	23,851,784

MATERIALS AND METHODS

The current investigation was performed to determine if lung CT images were benign or malignant.

Figure 2 shows a diagrammatic depiction of the proposed seven-stage technique. Stage I includes the collection of CT images of lungs from LIDC IDRI dataset. Stage-II includes image pre-processing. Stages III and IV consist of the study & implementation of pre-trained CNN models like VGG, Xception and MobileNet and calculate their performance by accuracy, precision, F1 score and recall. Study, implementation, and performance analysis of hybrid model VCNet is shown in stages

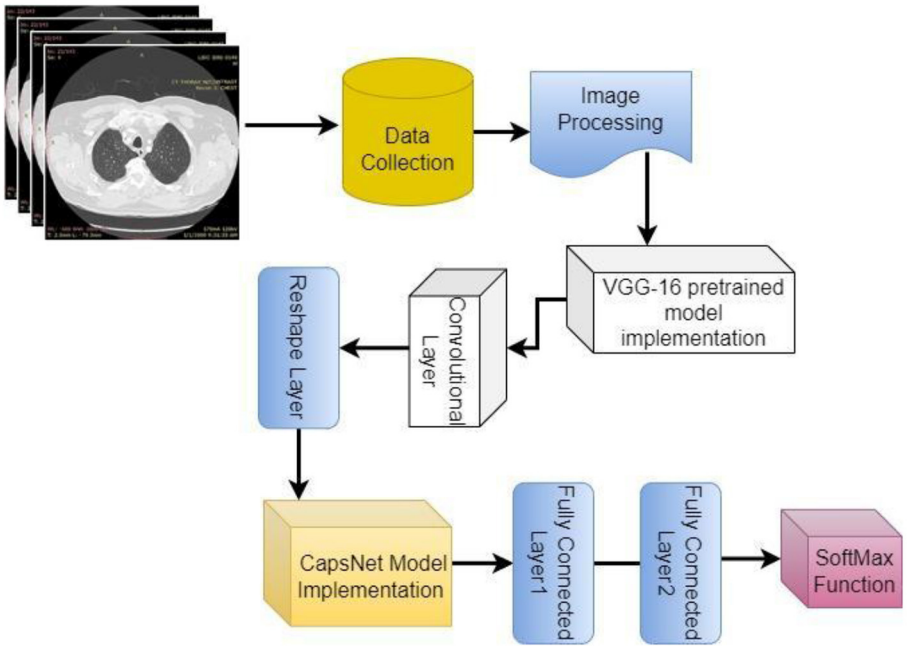


FIGURE 5 | Architecture of VCNet.

V & VI. Stage VII is the last phase, VCNet classifies CT scans as benign or malignant and analyzes their performance in terms of accuracy, precision, F1 score, and recall. All these stages are described in subsections.

Dataset Description

Dataset used here is LIDC- IDRI (11) that has a total 1,018 scans of 1,010 patients with 244,527 images. Each case in this dataset holds images of the CT scan and their corresponding XML file that contains annotations of the CT scan. Four experienced thoracic radiologists performed these annotations in two phases. In the first step, each radiologist individually categorized CT results into three classifications (nodules that are $\geq$3 mm, nodules that are $\leq$3 mm and non-nodule that are $\geq$3 mm). After that in the second step, classification is reviewed by each radiologist along with classification by another radiologist secretly. So every nodule of the dataset is assessed by all four radiologists separately. In this dataset, the diagnosis is done at two label patient levels and the nodule level. The DICOM images of the CT scan vary from 64 to 764 slices and have a resolution of 512*512*width. The average width of CT scans in this dataset is 240. Lung nodules in the LIDC-IDRI dataset are classified into four classes. (1) Unknown (No Label), (2) Benign or normal or non-cancerous (3) Primary lung cancer or Malignant, (4) Metastatic Lesion whose primary cancer was other than lung cancer.

Data Pre-processing

Preprocessing of the images is required to minimize the network overhead and computational complexity. The LIDC-IDRI dataset consists of various lung CT scan images and labeling is provided into csv file. For implementing the proposed model segregation and labeling of cancerous and noncancerous images are done based on annotations presented in csv file. Collected Dicom images are converted into jpg files using Radiant Dicom viewer 64 bit. Data augmentation methods used are rescaling, rotation, horizontal and vertical flip, and utilized ImageDataGenerator library of Keras for implementation. Rescaling is used to limit the image pixels between 0 to 1 and it helped in reducing computational complexity. Color Transformation of images (Gray to RGB) is done for better visualization. Rotation is done in a range of 10 means images are rotated from 0 to 10 degrees.

We have taken rotation range= 1 degree to randomly rotate our image between 0 and 1 degree. Rescale 1. /255 is done to transform every pixel value from range [0, 255] $\rightarrow$ [0, 1]. Color mode function of flowfrom directory of python is used to convert gray images into RGB images. For resizing the image from 512*512 to CNN input size, Keras ImageDataGenerator class is used that provides a quick and easy way to augment images.

The flow_from_directory () method of ImageDataGenerator is used to read the images directly from the directory and augment them while the neural network model is learning on the training data. The main important parameter of this method is target_size: Size of the input image. It is an attribute which converts images into size provided for the input. In our case we have taken image size 512*512, to set traget_size attribute value.

Convolutional Neural Network

Convolutional, pooling, and fully connected layers make up the three main layers of CNN design. For CNN classification and pattern recognition, the most popular architectures are LeNet, AlexNet, VGG-Net, Res-NET, ZF-Net, and SqueezNet.

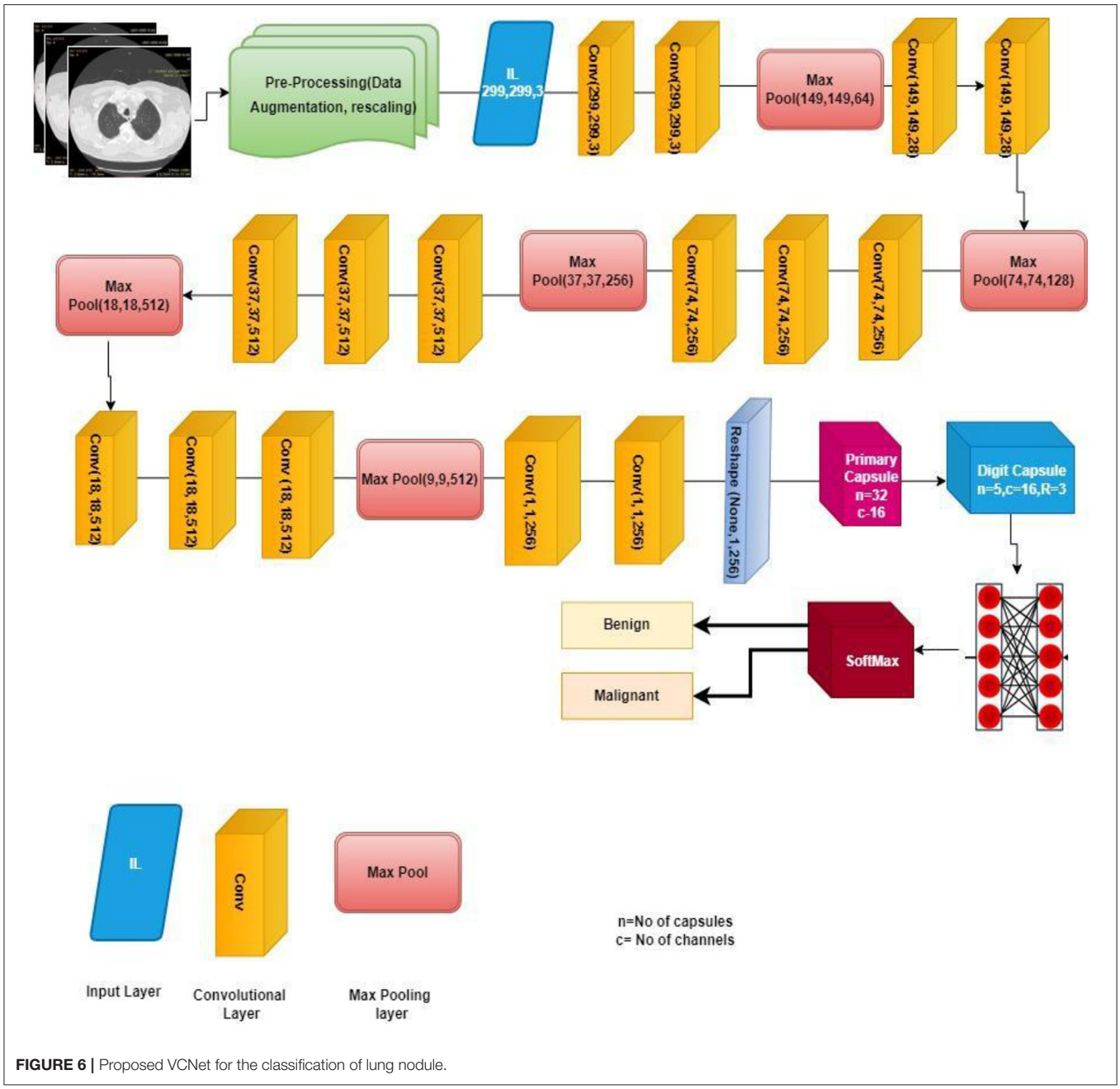


FIGURE 6 | Proposed VCNet for the classification of lung nodule.

Due to CNN's simplicity in calculation and function, the backpropagation algorithm is one of the finest learning algorithms. Back propagation differentiates during training between the algorithm predicting of labels and the ground truth labels computed by the loss function. Equation 1 gives the form of cross-entropy. Training samples are represented as inputs, and the true values of output neurons (j) are represented as inputs in the output layer (a_j^l) in this equation; n is the total number of training samples. It can be computed as:

$$C = -\frac{1}{n}\sum \times \sum_j \left(y_j \ a_j^l\right) \left(1 + y_j\right) l \, \mathrm{n}\left(1 - a_j^l\right). \quad (1)$$

Convolutional Layer

Each feature map comprises groups of neurons that together make up a feature map. Convolutional layer output is calculated by Eq. 2. Kernel (filter) size, feature maps, bias, and weight of a kernel are represented by F, m, B, and W_j, respectively. Convolutional layers have an output known as y_i, where I is an index that denotes the ith feature map in a layer known as l (2).

$$y_i^l = B_i^l + \sum_{j=1}^{ml(l-1)} F_{i,j}l \times W_j^{l-1} \quad (2)$$

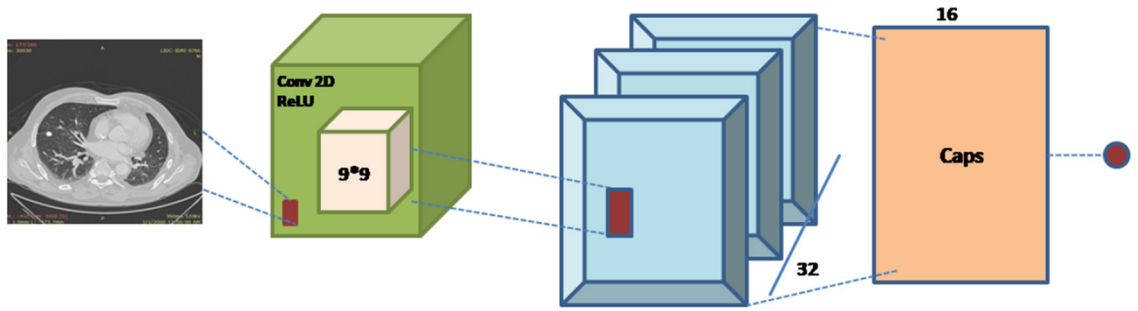


FIGURE 7 | CapsNet architecture.

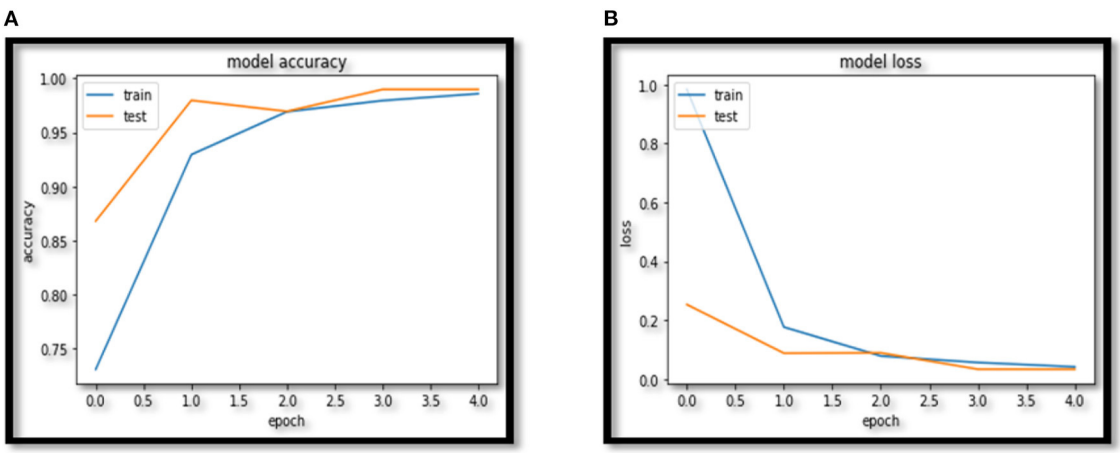


FIGURE 8 | (A) MobileNet accuracy. **(B)** MobileNet loss.

Pooling Layer

Generally, pooling layers are employed in collaboration with convolutional layers; as a result, sub-sampling decreases the input size in all depth portions and thus helps prevent over-fitting while training the network. The pooling process reduces the amount of the input, and as a result, the depth dimension is not altered. Maximum and average pooling is the most commonly used pooling technique, as shown in **Figure 3**.

There are two Equations in (3) that determine the breadth and height of the pooled-layer output.

$$B_2 = \left(\frac{B_1 + F}{S}\right) + 1 \tag{3}$$

$$H_2 = \left(\frac{H_1 + F}{S}\right) + 1 \tag{4}$$

Relu Activation

A nonlinear activation function is used in CNN instead of a local connection to determine the output of neurons. Using the improved performance, faster learning, and simple structure as a benefit, it is advantageous to use the logistic sigmoid and hyperbolic tangent activation functions (**Figure 4**). Equation (5) illustrates the Relu function. The gradient of the Relu function is zero if y is less than zero; otherwise, it is one (2).

$$f\left(y\right) = max(0, x) \tag{5}$$

Softmax Activation

Softmax is used to determine the probability of each ground truth label producing an output value between 0 and 1, then the results are transformed to perceptible values. The softmax function's formula is given by equation (6). $f\ (z)$ (22) transforms random variables (z) into meaningful values between 0 and 1 by using the softmax function.

$$f\left(z\right)_i = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \; for \; j = 1. \tag{6}$$

Performance Metrics

To examine the developed hybrid DL model for lung cancer nodule identification and classification using various performance metrics like accuracy, precision, recall, F1-score, and support, Overall accuracy is calculated as the total true results divided by the total number of samples. The blocks are

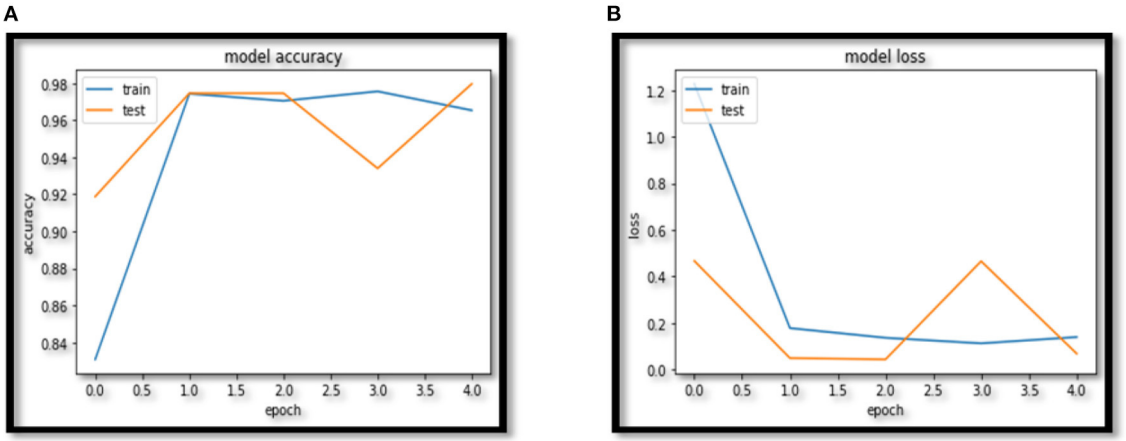


FIGURE 9 | (A) Xception accuracy. **(B)** Xception loss.

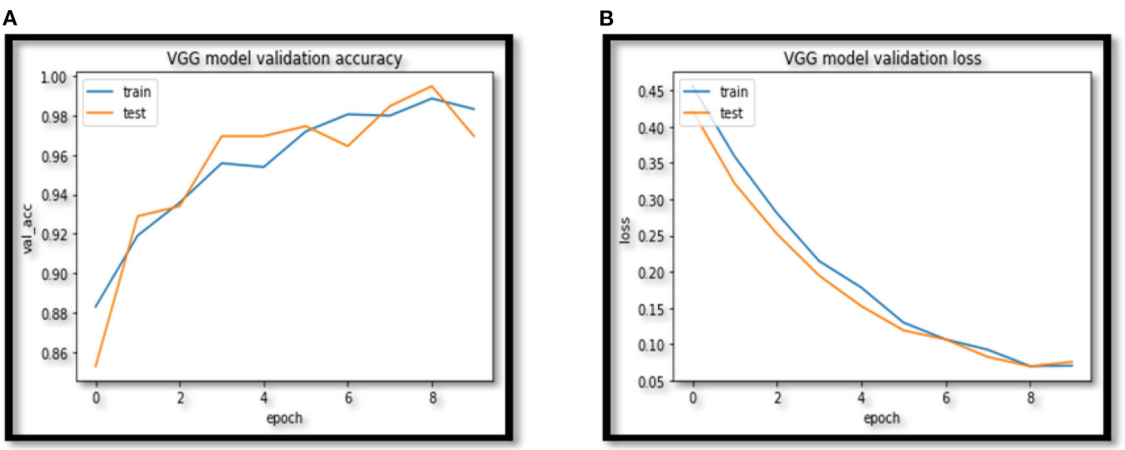


FIGURE 10 | (A) VGG accuracy. **(B)** VGG model loss.

derived from the pulmonary pictures in this study. These metrics are specified in several words like "True Positive" (TP) or "False Negative" (TN) and "False Negative" (FN). These are defined by different terminologies. TP is computed in this context as the suspected lung nodules that have been diagnosed properly as malignant. TN is determined by the number of pictures that are categorized as benign. The FN term is computed by the method that does not detect cancerous nodules. In addition, FP is determined by calculating the number of pictures not properly identified for lung cancer. Sensitivity and specificity are terms used to describe the true positive and true negative rates (23).The formulas for calculating these performance parameters are presented through Eqs 1–4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$F1\ Score = 2 \times \frac{Precision \times recall}{Precision + recall}. \tag{10}$$

Experimental Setup

The experiments are performed on CNN Pretrained models which are VGG-16, MobileNet and Xception on LIDC dataset. The new model VCNet is also developed and used for experimental purposes. The VCNet is designed by using a hybrid combination of VGG and Capsule Network.

Pretrained Models

We have used the following pre-trained models for the detection and classification of lung cancer using the LIDC-IDRI dataset.

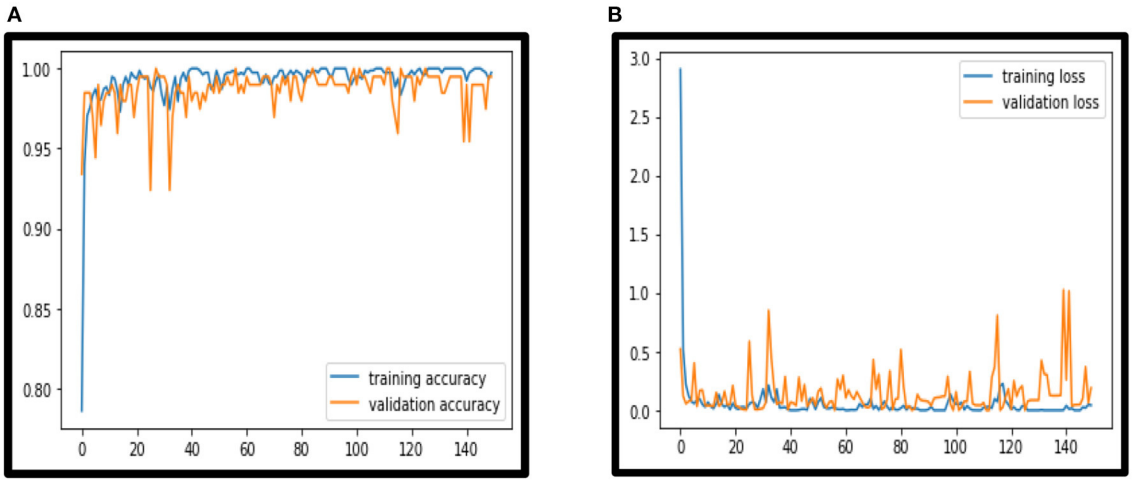


FIGURE 11 | (A) VCNet accuracy. **(B)** VCNet loss.

TABLE 3 | Comparative analyses in terms of accuracy among the proposed VCNET and the State-of-the-Art models.

Reference	Year	Image dataset	Accuracy
Jakimovski et al. (14)	2019	LONI Database	99.63%
Xie et al. (15)	2019	LUNA-16	86.42%
Qi et al. (16)	2020	Department of Radiology of the Henan Provincial People's Hospital	72%
Nakrani et al. (17)	2020	LIDC-IDRI	95.24%
Neal Joshua et al. (19)	2021	LUNA-16 dataset	97.17%
Afshar et al. (20)	2020	Chest X-ray dataset of NIH collected from kaggle	73%
Bharati et al. (21)	2020	LIDC-IDRI	83%
Proposed VCNet	**2022**	**LIDC-IDRI**	**99.49%**

Mobilenet

MobileNet employs depth-wise separable convolutions, which means that instead of combining all three and flattening, it executes a separate convolution on each color channel. The input channels are filtered as a result of this. For MobileNets, depth wise convolution applies a single filter to each input channel. The depth wise convolution's outputs are then combined using an 11 convolution by the point wise convolution. In one step, a conventional convolution filters and combines inputs to create a new set of outputs. The depth wise separable convolution divides this into two layers: one for filtering and the other for combining. This factorization results in a significant reduction in computation and model size (23).

Xception

Xception is a Depth wise Separable Convolutions-based deep convolutional neural network architecture. Google researchers came up with the idea. Inception modules in convolutional neural networks are described by Google as an intermediate step between normal convolution and the depth wise separable convolution operation (a depth wise convolution followed by a point wise convolution). In this sense, a depth wise separable convolution can be thought of as an Inception module with the most towers possible. This result leads them to propose new deep convolutional neural network architecture based on Inception, but with depth wise separable convolutions in place of Inception modules (24).

VGG-16

VGG16 as a convolutional neural network architecture in their study "Very Deep Convolutional Networks for Large Scale Image Recognition." On Image Net, a dataset of over 14 million images belonging to thousands of classifications, this model obtains 92.7 percent accuracy. It employs multiple 3 × 3 filters, one after the other. VGG16 was trained on the ImageNet dataset over weeks using the Nvidia Titan Black GPU. There are 13 convolution layers, five max-pooling layers, and three dense layers in the VGG16 pre-trained architecture. Changes to the VGG-16 algorithm include a global average pooling layer and two dense layers with the ReLu and Softmax activation functions. Both dense layers have a dropout rate of 0.5 (13).

Resnet

ResNet introduced by Kaiming He in 2015, it has become key work proving that incredibly deep networks may be trained using regular SGD with additional accuracy by changing the residual module to incorporate identity mapping, as proven in this study. This network employs a VGG-19-inspired 34-layer plain network architecture, after which the shortcut connection is implemented. The design is subsequently transformed into a residual network as a result of these shortcut connections (25).

Inception V3

Inception-v3 is a convolutional neural network design from the Inception family that includes Label Smoothing, Factorized 7

× 7 convolutions, and the inclusion of an auxiliary classifer to transport label information deeper down the network, among other improvements. Inception v3 is a commonly used image recognition model that has been demonstrated to achieve an accuracy of better than 78.1 percent on the ImageNet dataset. The model represents the sum of several concepts produced over the years by numerous scholars (26).

Table 2 presents the architecture details and number of parameters used for MobileNet, Xception, and VGG-16, ResNet & Inception V3.

VCNeT Model Building and Implementation

The proposed hybrid model is implemented on the Google Colab platform by utilizing the python libraries Tensorflow and Keras. For the transfer learning pre-trained model of CNN, VGG-16 is used and modified by adding global average pooling layers and dense layers with ReLu activation function. After the VGG, a capsule network has been added with route 03. Our results show that the hybrid model VCNet performs better than the VGG model, Xception, and MobileNet architectures.

Figure 5 shows the complete architecture of the VCNet and summary of the proposed model is shown in **Figure 6**.

Implementation of Capsule Network

For the implementation of capsule network Global_Average_Pooling layer is used as an input and after that Convolution 2-dimensional layer is used to extract features with a kernel size of 9,9,512 and these features are mapped using reshape function to convert the images into one-dimensional array using reshape (1,1,3) along with squash function which is used for normalization. Lambda ?? is used after this to show that the lung image characteristics had a normal value of 0 with the default route transfer rate [−0.5: 0.5]. Calculating a margin and reconstruction loss is the most recent loss computation. This article uses the CapsNet for lung imaging data from Hinton's major architecture (26). **Figure 7** shows a basic architecture of capsule network (CapsNet) for analysis of lung nodules.

The capsule network architecture is summarized as follows:

The first Convolution layer with kernel_size = 9, filters = 512, padding = "same," strides = 9, activation = "relu," dim_capsule = 8, *n*_channels = 32, padding = "valid," strides = 2, kernel_size = 9 for the second primary capsule. *n*_class = num_capsule, dim_capsule = 16, stable of the set routings for the diagnosis capsule.

RESULT

In this paper VCNet, a new hybrid DL model is proposed for the detection of lung carcinoma using CT images. The proposed model is applied to CT images collected from the LIDC-IDRI dataset. The accuracy presented by VCNet architecture is 99.49% which is the highest among other used pretrained models. The pretrained models MobileNet, Xception, and VGG-16 achieved accuracy of 98, 97.97 and 96.95% respectively. The hybrid model built is evaluated with a validation set of pulmonary pictures. The accuracy and loss graphs of MobileNet architecture are illustrated in **Figures 8A,B**, **9A,B**, Xception architecture accuracy and loss graph, **Figures 10A,B**, VGG-16 architecture accuracy and loss graph and **Figures 11A,B** represent the accuracy and loss graph for VCNet's proposed architecture.

The findings in **Table 3** shows that suggested hybrid model VCNet were more accurate than other models in classifying lung CT images. As a consequence, the suggested hybrid VCNet attained the highest accuracy, which was 99.49 percent.

Classification matrix of different models for detection of lung nodules using LIDC IDRI dataset is presented in **Table 4**.

The total number of properly identified lung nodules and their related erroneous predictions are shown in the confusion matrix created after evaluation in **Figures 12A–D**. For the statistical analysis the performance of the hybrid model VCNet is compared with various competitive models such as CNN, MobileNet, Xception, VGG-16, ResNet and InceptionV3.

DISCUSSION

There are numerous DL models for lung cancer diagnosis; nevertheless, accuracy in public datasets is still a significant obstacle (27). Overall, the researchers' results were accurate, although some of them utilized either private hospital datasets or tiny datasets that were publicly available on the internet. 95% accuracy was obtained by comparing two big artificial neural network end-to-end classes, one of which was the last stage of convolutional neural network training and the other the final stage of artificial neural network training. The implementation of the same employed the unpublished dataset of 31 patients that had 38 scans (13).

TABLE 4 | Classification matrix of different models using LIDC dataset.

Model	Accuracy	F-Measure	Sensitivity	Specificity	AUC
CNN	95.88 ± 2.01	95 ± 1.84	96 ± 1.85	96 ± 1.92	96 ± 1.02
Xception	97.97 ± 1.02	96.87 ± 1.82	97.47 ± 1.05	97.97 ± 1.02	96.97 ± 1.02
VGG-16	96.95 ± 1.75	96.95 ± 1.55	95.95 ± 1.75	96.95 ± 1.25	95.95 ± 1.84
MobileNet	98 ± 0.40	98 ± 0.55	98 ± 0.65	98 ± 0.40	98 ± 0.35
ResNet	93.5 ± 1.38	93.5 ± 1.55	93.5 ± 1.28	93.5 ± 1.4	93.5 ± 1.38
Inception V3	94.35 ± 1.1	94.35 ± 1.15	94.35 ± 1.2	94.35 ± 1.10	94.35 ± 1.1
VCNet	**99.49 ± 0.2**	**99.12 ± 0.5**	**99.25 ± 0.5**	**99.25 ± 0.6**	**99.12 ± 0.6**

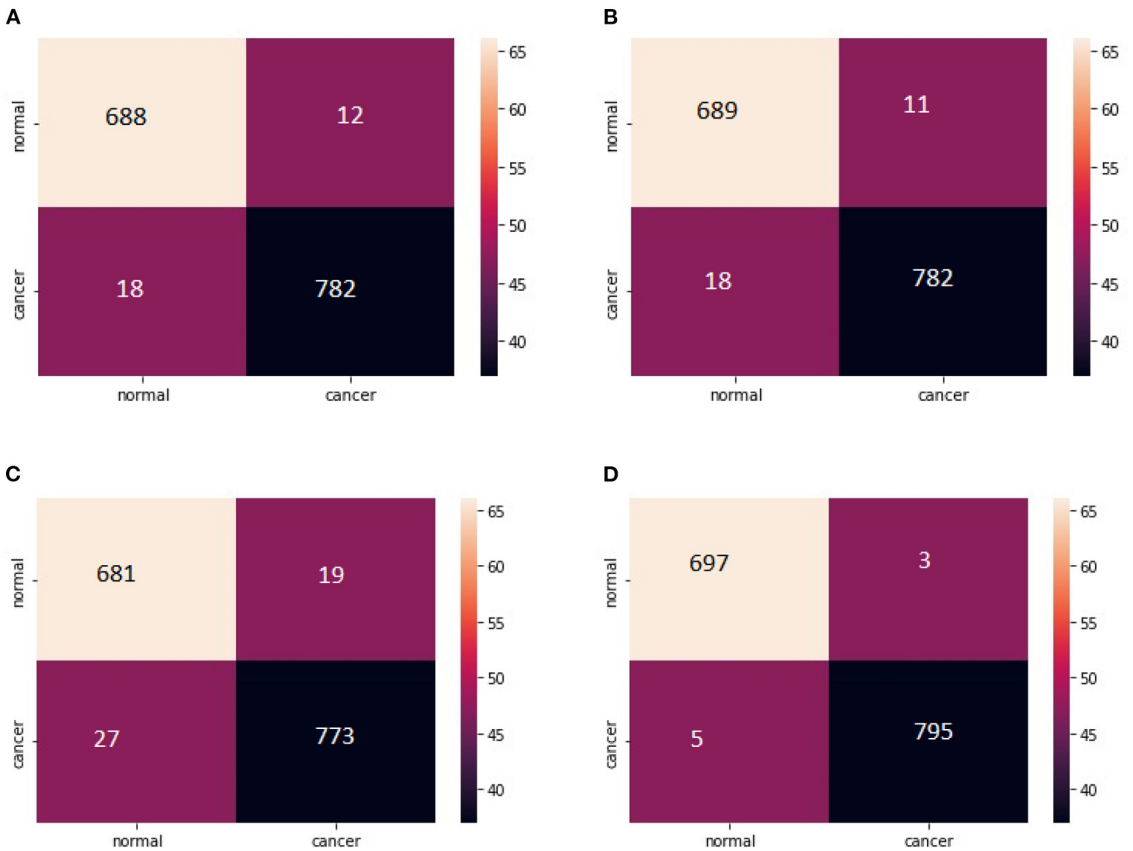


FIGURE 12 | (A) Confusion Matrix MobileNet. **(B)** Confusion Matrix XceptionNet. **(C)** Confusion matrix VGG-16. **(D)** Confusion matrix VCNet.

Detecting malignant lung nodules from the CT images takes a significant amount of time, especially if you don't have prior training in the subject. Despite recent progress in using deep learning models for lung carcinoma detection, clinical use is still difficult. For the majority of the cases, deep learning neural network models are used to identify the main nodule of the lung while ignoring the tissue that surrounds it.

In this section, the output of VCNet architecture and pre-trained models of CNN VGG-16, MobileNet and Xception on collected lung CT images are presented for classification of lung cancer nodules as normal and malignant. There were no issues applying the models because of the LIDC-IDRI dataset after preprocessing.

According to this study, 80 % of said dataset is used as a training set and the remaining 20 percent is used to evaluate the suggested VCNet architectures and pre-trained VGG-16. MobileNet& Xception architectures. **Table 4** provides a summary of the Accuracy, Precision, Recall, and F1 scores for each of the four architectures.

The model's validation accuracy and loss curves are shown in **Figures 8–11**. Using the graphs, we may conclude that no over fitting or under fitting has occurred. Training and validation accuracy and losses continue to converge, and they reach a maximum at the end of 30 epochs. On the proposed VCNet, examined in terms of validation accuracy, it was determined that the network has an error rate of $<0.1\%$ and a test accuracy of $<0.1\%$,demonstrating a good correlation between the actual and predicted values.

For lung nodules, the estimated values of Specificity and Sensitivity are near to 1 shown in confusion matrix, indicating that the model has a high true positive and true negative rate. Furthermore, the low false positive and negative rates indicate that the proposed VCNet model has a very low risk of misclassifying data. As a result, the algorithm was able to properly detect lung nodules with extremely little error.

The proposed hybrid model evaluated on the Lung Picture Database Consortium image collection dataset achieved higher accuracy as compared to current techniques MobileNet, Xception, and VGG-16. It is critical for any system in the medical sector to produce accurate findings with a little mistake, since this framework allows for clinical purposes; it can be utilized for the detection of lung nodules in lung carcinoma.

CONCLUSION

As a result of deep learning methods' impressive outcomes, in this research, CNN-based hybrid architecture, one of the effective models of deep learning in medical research, was

applied to examine lung cancer on CT scan pictures of the LIDC-IDRI dataset. Therefore, prior to implementing the model, a collection of CT scan pictures was generated from an image dataset.The proposed hybrid model is applied to CT images collected from the LIDC-IDRI dataset. The proposed VCNet architecture proposed a higher accuracy of 99.49% while the existing methods of MobileNet, Xception, and VGG-16 achieved an accuracy of 98, 97.97, and 96.95%, respectively. This proposed hybrid VCNet framework can be used for clinical purposes for nodule detection in lung carcinoma detection, which will be a very useful and time-saving method for radiologists.

Future Scope

In the future this hybrid model can be used for the detection of other types of cancer like Breast, Liver, Cervical, Brain, and Skin etc. The performance of the model can be optimized by using various optimization techniques also.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

RT: review of literature, methodology, experiments and result preparation, and draft preparation of paper. SA: methodology, proof reading of paper, and experimental design. AC: review of literature, result preparation and funding support. SB: proof reading of paper and funding support. All authors contributed to the article and approved the submitted version.

REFERENCES

1. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
2. Agrawal M, Agrawal S. Rice plant diseases detection and classification using deep learning models: a systematic review. *J Crit Rev*. (2020) 7:4376−90. doi: 10.31838/jcr.07.11.592
3. Agrawal S, Jain SK. Medical text and image processing: applications, issues and challenges. In: Jain V, Chatterjee J, editors. *Machine Learning with Health Care Perspective*. Cham: Springer (2020). p. 237–62. doi: 10.1007/978-3-030-40850-3_11
4. Tandon R, Agaral S, Goyal P. Sequential Convolutional Neural Network for Automatic Breast cancer image classification using Histopathological Images. *Methods*. (2020) 7:10. doi: 10.31838/jcr.07.15.779
5. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
6. Kutia S, Chauhdary SH, Iwendi C, Liu L, Yong W, Bashir AK. Socio-Technological factors affecting user's adoption of eHealth functionalities: a case study of China and Ukraine eHealth systems. *IEEE Access*. (2019) 7:90777–88. doi: 10.1109/ACCESS.2019.2924584
7. Dargan S, Kumar M, Rohit Ayyagari M, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. *Arch Comput Methods Eng*. (2020) 27:1–22. doi: 10.1007/s11831-019-09344-w
8. Saravagi D. Agrawal S. Opportunities and challenges of ML model for prediction and diagnosis of spondylolithesis: a systematic review. *Int J Eng Syst Model Simul*. (2021) 12:127–38. doi: 10.1504/IJESMS.2021.115534
9. Iwendi C, Srivastava G, Khan S, Maddikunta PKR. Cyberbullying detection solutions based on deep learning architectures. *Multimed Syst*. (2020) 1–14. doi: 10.1007/s00530-020-00701-5
10. Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. In: *Advances in Neural Information Processing Systems*. (2017), p. 3856–66.
11. Armato Samuel G III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. *Data From LIDC-IDRI*. (2015). Available online at: https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI (accessed January 7, 2020).
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. (2014). doi: 10.48550/arXiv.1409.1556
13. Tajbakhsh N, Suzuki K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. *CNNs Pattern Recognit*. (2017) 63:476–86. doi: 10.1016/j.patcog.2016.09.029
14. Jakimovski G, Davcev D. Using double convolution neural network for lung cancer stage detection. *Appl Sci*. (2019) 9.3:427. doi: 10.3390/app9030427
15. Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognit*. (2019) 85:109–19. doi: 10.1016/j.patcog.2018.07.031
16. Qin R, Wang Z, Jiang L, Qiao K, Hai J, Chen J, et al. Fine-grained lung cancer classification from PET and CT images based on multidimensional attention mechanism. *Complexity*. (2020) 2020:165–75. doi: 10.1155/2020/6153657
17. Nakrani MG, Sable GS, Shinde UB. ResNet based lung nodules detection from computed tomography images. *Int J Innov Technol Exploring Eng*. (2020) 1711–4. doi: 10.35940/ijitee.D1540.029420
18. Lu X, Nanehkaran YA, Fard MK. A method for optimal detection of lung cancer based on deep learning optimized by marine Predators algorithm. *Comput Intell Neurosci*. (2021). doi: 10.1155/2021/3694723. [Epub ahead of print].
19. Neal Joshua ES, Bhattacharyya D, Chakkravarthy M, Byun YC. 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. *J Healthcare Eng*. (2021). doi: 10.1155/2021/6695518. [Epub ahead of print].
20. Afshar P, Oikonomou A, Naderkhani F, Tyrrell PN, Plataniotis KN, Farahani K, et al. 3D-MCN: a 3D multi-scale capsule network for lung nodule malignancy prediction. *Sci Rep*. (2020) 10.1:1–11. doi: 10.1038/s41598-020-64824-5
21. Bharati S, Podder P, Mondal MRH. Hybrid deep learning for detecting lung diseases from X-ray images. *Inform Med Unlocked*. (2020) 20:100391. doi: 10.1016/j.imu.2020.100391
22. Bouchard G. Efficient bounds for the softmax function and applications to approximate inference in hybrid models. In: *Proceedings of the Presentation at the Workshop For Approximate Bayesian Inference in Continuous/Hybrid Systems at Neural Information Processing Systems (NIPS)*, Meylan, France Vol 31. (2008).
23. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv*. (2017). doi: 10.48550/arXiv.1704.04861
24. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2017). doi: 10.1109/CVPR.2017.195
25. Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv*. (2020). doi: 10.48550/arXiv.2010.16061
26. Sadiq MT Yu X, Yuan Z. Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces. *Expert Syst Appl*. (2021) 164:114031. doi: 10.1016/j.eswa.2020.114031

27. Sadiq MT, Yu X, Yuan Z, Aziz MZ. Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform. *Electron Lett*. (2020) 56:1367–9. doi: 10.1049/el.2020.2509

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tandon, Agrawal, Chang and Band. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 22 June 2022
doi: 10.3389/fpubh.2022.926069

XGBoost-Based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers

*Xianbin Song[1], Jiangang Zhu[1], Xiaoli Tan[2], Wenlong Yu[1], Qianqian Wang[1], Dongfeng Shen[1] and Wenyu Chen[2]**

[1] Department of Critical Care Medicine, Affiliated Hospital of Jiaxing University, Jiaxing, China, [2] Department of Respiration, Affiliated Hospital of Jiaxing University, Jiaxing, China

In December 2019, an outbreak of novel coronavirus pneumonia spread over Wuhan, Hubei Province, China, which then developed into a significant global health public event, giving rise to substantial economic losses. We downloaded throat swab expression profiling data of COVID-19 positive and negative patients from the Gene Expression Omnibus (GEO) database to mine novel diagnostic biomarkers. XGBoost was used to construct the model and select feature genes. Subsequently, we constructed COVID-19 classifiers such as MARS, KNN, SVM, MIL, and RF using machine learning methods. We selected the KNN classifier with the optimal MCC value from these classifiers using the IFS method to identify 24 feature genes. Finally, we used principal component analysis to classify the samples and found that the 24 feature genes could effectively be used to classify COVID-19-positive and negative patients. Additionally, we analyzed the possible biological functions and signaling pathways in which the 24 feature genes were involved by GO and KEGG enrichment analyses. The results demonstrated that these feature genes were primarily enriched in biological functions such as viral transcription and viral gene expression and pathways such as Coronavirus disease-COVID-19. In summary, the 24 feature genes we identified were highly effective in classifying COVID-19 positive and negative patients, which could serve as novel markers for COVID-19.

Keywords: COVID-19, diagnostic markers, XGBoost, machine learning, principal component analysis

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Vinit Gunjan, CMR Institute of Technology, India
B. V. Santhosh Krishna, New Horizon College of Engineering (NHCE), India

***Correspondence:**
Wenyu Chen
00135116@zjxu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *22 April 2022*
Accepted: *31 May 2022*
Published: *22 June 2022*

Citation:
Song X, Zhu J, Tan X, Yu W, Wang Q, Shen D and Chen W (2022) XGBoost-Based Feature Learning Method for Mining COVID-19 Novel Diagnostic Markers. Front. Public Health 10:926069. doi: 10.3389/fpubh.2022.926069

INTRODUCTION

In December 2019, an epidemic of novel coronary pneumonia broke out in Wuhan, Hubei Province, China, which was considered by the World Health Organization to be a serious menace to the health of citizens of the world (1). This terrible communicable epidemic is caused by infection with the severe acute respiratory syndrome type 2 coronavirus (SARS-CoV-2), a sense single-stranded RNA virus (2). As a highly contagious virus, COVID-19 swept across the globe with alarming rapidity, leading to considerable losses to human society.

So far, the effective protection strategy against COVID-19 is to strengthen immunity ability and keep social distance (3). COVID-19 diagnosis is of great essence for the identification, isolation, and treatment of infectious objects (4). Existing detection methods include antibody assays that detect serum antiviral antibodies IgG and IgM, lateral chromatography assays that detect viral antigens, and real-time reverse transcriptase-polymerase chain reaction (qRT-PCR). The current gold standard for COVID-19 diagnosis is the application of qRT-PCR to verify the presence of

SARS-CoV-2 RNA in the respiratory secretions of patients (5, 6). However, this detection method is not perfect because it is a complex test requiring a comprehensive and delicate infrastructure (5). And this method can only achieve accuracies of 30–60% in clinical application, which probably results in false-positive cases (7). More landmark diagnostic biomarkers are needed to detect COVID-19-positive patients with higher accuracy, reducing the false positive rate. Besides, exploring and developing new detection kits is of equal significance to facilitate the precise prevention and control of the epidemic.

Machine learning is applied extensively in biomedical applications, as well as COVID-19 diagnosis (8). Extreme Gradient Boosting (XGBoost) is a GBDT-based algorithm. Characterized by its high efficiency, flexibility, and portability, XGBoost is widely used in data mining, recommendation systems, and other fields (9). Zhang and GuoLiang (10) developed a machine learning algorithm for XPPA based on the XGBoost algorithm, which could be used to detect the effect of alterations in gene expression on aberrant p53 pathway activity. Athanasiou et al. (11) constructed a personalized risk prediction model for cardiovascular disease based on the XGBoost algorithm to predict the incidence of patients with cardiovascular disease. The follow-up results of 560 patients demonstrate that this predictive model has favorable performance (AUC = 71.13%), which is expected to provide new insights into clinical cardiovascular treatment. With a decoupling feature, XGBoost shows increased applicability, and it is a high-performance algorithm for modeling regarding the selection of loss functions on demand for classification and regression. Therefore, XGBoost is reliable to be applied in establishing a diagnostic, prognostic model based on patient features in clinical practice.

Here, we used the XGBoost algorithm to mine feature genes in the expression profiles of COVID-19 negative and positive samples, used a machine learning algorithm to construct MARS, KNN, SVM, MIL, and RF COVID-19 classifiers, and selected the best classifier using Iterated Function System (IFS) algorithm. Finally, the validity of this set of feature genes was verified by principal component analysis (PCA) and functional enrichment analysis, the results of which suggested the potential of the genes to be promising biomarkers for COVID-19.

MATERIALS AND METHODS

Datasets Downloading and Processing

From the GEO database (https://www.ncbi.nlm.nih.gov/geo/), the dataset GSE152075 was downloaded, which contained gene expression data from throat swab samples from 430 COVID-19-positive patients and 54 negative patients. And the data acquisition platform was GPL18573 (Illumina NextSeq 500). Genes whose mean value of gene expression was below 1 and the maximum value of gene expression was below 5 were retained. The data were normalized using the "edgR" package (12).

Model Training

To establish the link between behavioral features and classification, we implemented the XGBoost model using the machine learning algorithm XGBoost (https://xgboost.ai/). Key features were determined based on feature importance ranking and recursive elimination (9). XGBoost is a gradient advancing decision tree method whose objective function is defined as in Equation (1).

$$\$(\phi) = \sum_{i=1}^{n} loss\left(y_{i\hat{y}_i}\right) + \sum_{k=1}^{k} \Omega\left(f_k\right)(1) \quad (1)$$

In this formula, loss is the training loss, Ω (f) is the complexity of the tree, and k is the number of trees in the model. The model can be optimized by minimizing the objective function. For this reason, the additive model was used to calculate the training loss, and the Taylor expansion method was used to quickly optimize the prediction of the nth round of additive training. Greedy algorithm was used to determine the optimal complexity of the tree. In addition, we employed SMOTE for Bayesian optimization resampling of the training set due to unbalanced samples (13).

Selecting the Optimal Classifier by IFS Method

After feature selection by XGBoost, IFS method was used to identify the genes of the optimal COVID-19 classifier. IFS incremental feature selection method (14) is an algorithm proposed by Liu and Setiono (15) to find the best or closest optimal feature subset. This algorithm is based on improved information gain, which can make the equivalent exchange of information. The algorithm selects a candidate feature set using an evaluation function unrelated to the classifier, applies the classifier to the candidate feature set, and selects a feature subset utilizing the accuracy of the classifier as a criterion.

A series of COVID-19 classifiers (16) was subsequently established using the python package "sklearn" in combination with algorithms such as MARS, KNN, SVM, MIL, and RF. The IFS curve was drawn based on 10-fold cross-validation, resulting in Matthews correlation coefficient (MCC) for each classifier, which is a parameter that can effectively reflect the classifier's effectiveness (17). The classifier with the most considerable MCC value is considered as the optimal classifier, and the genes involved in it are taken as the optimal feature genes.

PCA and Sample Cluster Analysis

After the optimal COVID-19 classifier was determined, the PCA was performed on the data set using "FactoMineR" to extract the first and second principal components. PCA analysis is an unsupervised dimensionality reduction analysis method which can visually present the sample-to-sample method (18) by reducing the dimensionality of the dataset and reflecting the data to the representative dimensions PC_1 and PC_2. The effect of model classification was finally verified by pedigree cluster analysis of the samples using the "pheatmap" package (19).

GO and KEGG Enrichment Analyses

GO biological function analysis and KEGG biological pathway analysis of feature genes were performed using "clusterProfiler". GO and KEGG pathways with p-value < 0.05 were considered notably enriched (20).

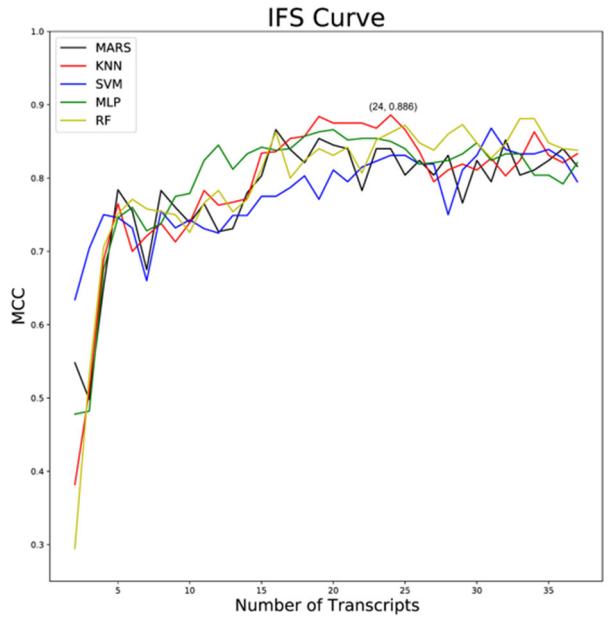


FIGURE 1 | IFS curves of MARS, KNN, SVM, MLP, and RF classifiers. Black: MARS classifier; red: KNN classifier; blue: SVM classifier; green: MLP classifier; brownish-yellow: RF classifier; horizontal ordinate indicates the number of classifier genes and vertical ordinate represents MCC coefficient.

RESULTS

XGBoost and IFS Analysis Results

A total of 15,190 genes were obtained by the normalization of the gene expression data after the preprocessing of the dataset GSE152075. Thirty-seven feature genes that were ranked according to importance were obtained by XGBoost feature selection, which could distinguish sample types (**Supplementary Table S1**). And COVID-19 classifiers such as MARS, KNN, SVM, MLP, and RF were constructed based on these 37 feature genes. Then the best classifier was selected by IFS analysis, and we found that the KNN classifier was composed of 24 feature genes (IGFBP2, KRT8, RPLP0, XAF1, RPL13, OAS2, CES1, RPL4, EEF1G, NR2F6, RPS8, RPL10A, SNX14, C5orf15, TNFRSF19, CD24, ALAS1, CEP112, C9orf24, POLR2J3, AAMP, DUOX2, EMCN, RPL3) had the highest MCC value, MCC 0.886, sensitivity 0.986, specificity 0.907, and accuracy 0.977 (**Figure 1**).

The Results of PCA Dimensionality Reduction Analysis and Sample Cluster Analysis

PCA dimensionality reduction analysis was performed on the samples according to the expression of the 24-feature genes in the optimal KNN classifier, which showed that PCA analysis could classify COVID-19 in positive patients and negative persons (**Figure 2A**). In addition, we also plotted a cluster heatmap analyzing the expression of 24 feature genes in different populations. The results showed that the 24 feature genes in the KNN classifier could distinguish COVID-19 positive patients from normal healthy people (**Figure 2B**). These findings indicated that the 24 feature genes in the KNN classifier performed well in diagnosing COVID-19-positive patients and normal healthy people, showing superior diagnostic efficacy.

The Results of GO and KEGG Enrichment Analyses

To identify the biological functions of feature genes and the signaling pathways involved, we performed enrichment analyses on the 24 feature genes. The GO analysis result showed that these genes were mainly enriched in biological functions such as viral transcription and viral gene expression (**Figure 3A**). KEGG biological pathway analysis showed gene enrichment on pathways such as Coronavirus disease-COVID-19 (**Figure 3B**). The selected feature genes were closely related to COVID-19 infection and its pathways.

DISCUSSION

Novel coronavirus pneumonia is a severe threat to global public health safety and brings enormous economic losses to human society. In this study, in order to identify new COVID-19 diagnostic biomarkers, we used the XGBoost algorithm to achieve feature selection and the IFS algorithm to determine the optimal classifier based on the throat swab expression profile data of COVID-19 positive and negative samples in the GEO database. After identifying the optimal feature genes, PCA, GO, and KEGG methods were used to verify whether the feature genes could be used as COVID-19 diagnostic biomarkers. First, we used the XGBoost algorithm to screen 37 feature genes from expression profiling data that could effectively distinguish COVID-19 positive from negative patients. Subsequently, KNN, SVM, MLP, and RF classifiers were constructed for the genes after feature selection, and the optimal classifier and its feature genes were selected based on the IFS method. Finally, we identified 24 feature genes, and based on the expression data of 24 feature genes, we performed PCA of the samples, and PCA results showed that PC_1 and PC_2 could effectively distinguish COVID-19 positive and negative samples. In addition, we performed GO and KEGG enrichment analyses of 24 feature genes, and the results showed that these feature genes were mainly gathered in biological functions such as viral transcription, viral gene expression, and pathways such as Coronavirusdisease-COVID-19. Therefore, combining all the results of bioinformatics analysis, the COVID-19 classifier of 24 feature genes was obtained in this study, while we reasonably speculated that the 24 feature genes screened in this study are expected to be novel diagnostic biomarkers for COVID-19.

Timely diagnosis of COVID-19 is essential for epidemic prevention and control, so identification of accurate diagnostic biomarkers is also an essential study for epidemic prevention and control. Feng et al. (21) constructed a machine learning diagnostic model using algorithms such as LASSO, AdaBoost, decision tree, and logistic regression based on patient clinical information to assist early COVID-19 diagnosis. The study by Kukar et al. (22) used machine learning methods to construct a

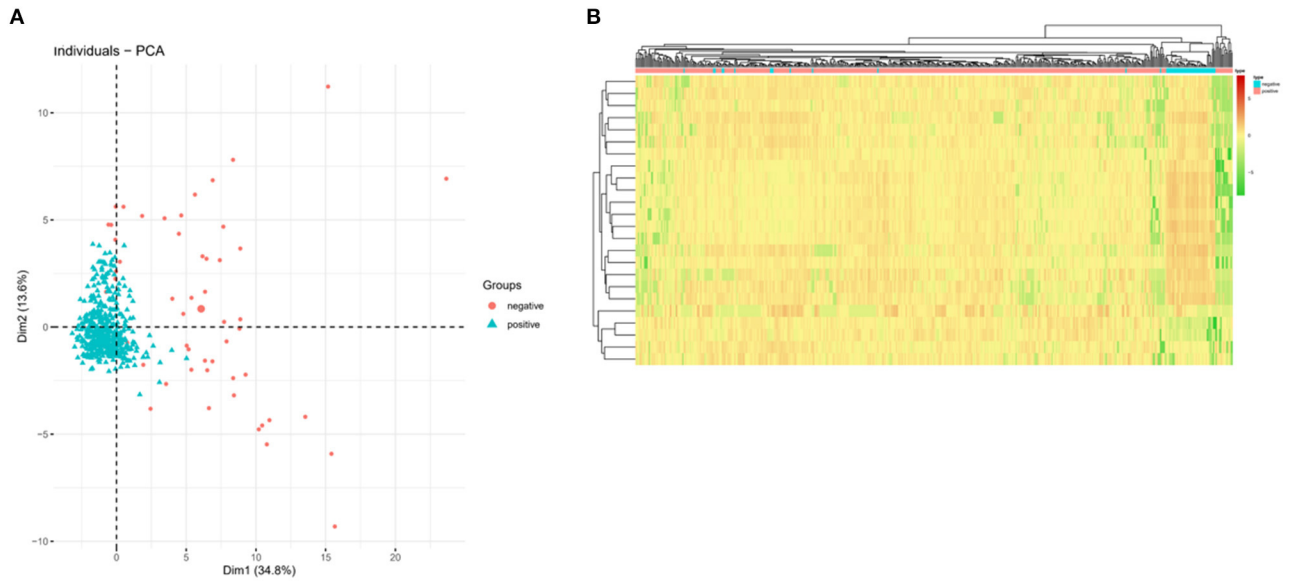


FIGURE 2 | PCA and cluster heatmap analysis based on feature genes in the KNN classifier. **(A)** PCA shows the classification performance of the KNN classifier in COVID-19 negative (red) and positive (green) populations. **(B)** Cluster heatmap showing the expression of feature genes in the KNN classifier. Red indicates high expression and green indicates low expression.

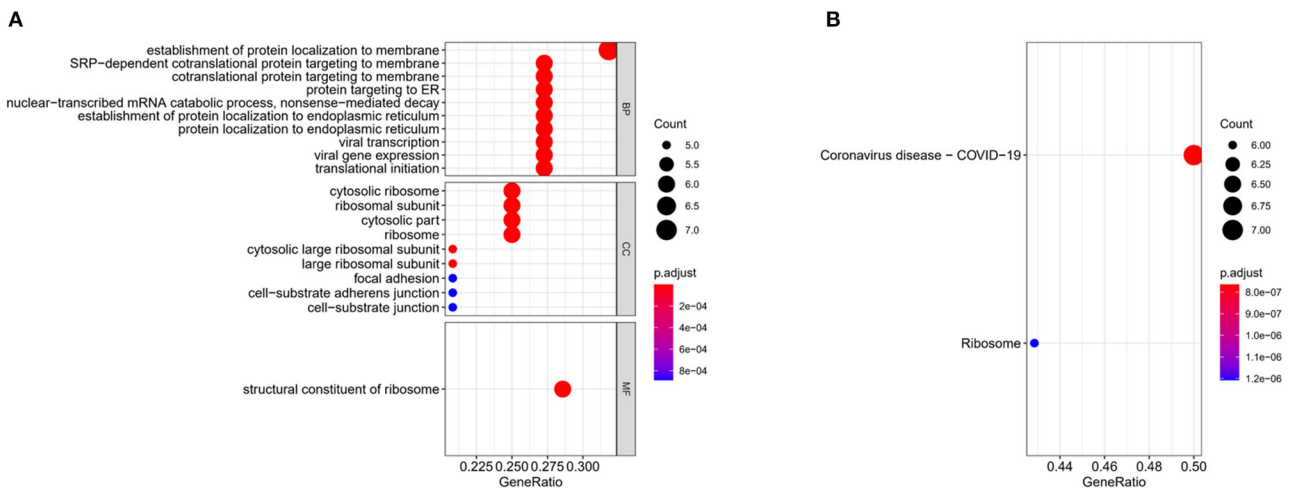


FIGURE 3 | Gene enrichment analyses. **(A)** Bubble plots for GO enrichment analysis of 24 feature genes. **(B)** Bubble plots for KEGG enrichment analysis of 24 feature genes. The bubble size in the figure indicates the gene data in teams, and the color indicates the *p*-value, and the red the color, the smaller the *p*-value.

COVID-19 diagnostic model based on blood routine parameters, which is complementary to chest CT and PT-PCR molecular diagnostics and improves COVID-19 diagnostic efficiency. Our study used the XGBoost algorithm to select feature genes in the expression profiles of throat swabs in positive patients, constructed classifiers such as MARS, KNN, SVM, MIL, and RF, and subsequently selected classifiers with optimal MCC values by the IFS method. At present, the conventional detection method of COVID-19 is nucleic acid detection, and the diagnostic biomarkers identified in this study are expected to improve the drawbacks of existing commercial nucleic acid detection kits and improve detection accuracy.

The optimal 24 feature genes, which were further analyzed by consulting the retrieved literature, we found that four genes (XAF1, OAS2, CES1, RPS8) have been reported in COVID-19. Gao et al. (23) found that XAF1 was abnormally strongly expressed in COVID-19 patients and positively correlated with the expression of ARS-CoV-2 invasion-related genes (ACE2, TMPRSS2, CTSB, and CTSL). In contrast, XAF1 was found to be associated with SARS infection by Park and Harris (24). A

recent study found that OAS2 belongs to a subset of interferon-stimulated genes, and OAS2 can be regarded as a potential candidate for a drug target in COVID-19 therapy (25). The study by Li et al. (26) found that CES1 can hydrolyze tenofovir alafenamide (TAF), and effectively hydrolyzed TAF is significant for treating respiratory virus infection. In addition, Vastrad et al. (27) identified 10 SARS-CoV-2/COVID-19 diagnostic markers such as RPS8 using bioinformatics analysis methods. Also, several ribosomal proteins (RPL family members) contributing to protein synthesis were screened out. A report went that SARS-CoV-2 infection could result in ribosome dysfunction (28), giving us a hint that RPLs were affected at molecular degree. In combination with previous reports, it can be seen that some of the 24 feature genes are closely related to COVID-19. Finally, we performed GO and KEGG enrichment analyses, and the results showed that these feature genes were mainly enriched in biological functions such as viral transcription and viral gene expression as well as pathways such as Coronavirusdisease-COVID-19. We used bioinformatics methods to screen some genes that play an essential role in COVID-19 infection, which have also been reported as COVID-19-related genes in the existing literature. Even though it takes little time and hardly any money to detect COVID-19, some critical problems remain, like false positive case which concerns the public a lot. The combined various testing methods are urgently needed to remove false positive cases. Our study comes just in handy to provide some insights for developing novel strategy for COVID-19 diagnosis, which can definitely enrich current diagnostic tools.

CONCLUSION

However, there are limitation in our study. First, this study is a retrospective study based on public databases, and no clinical samples are used to verify the performance of this classifier. Second, even if the mined genes were practically used for COVID-19 diagnosis, it is relatively costing to analyze 24 genes for one sample. Considering the limitations, we are planning to establish sample library and validate our model based on our collected samples. Overall, we mined optimal COVID-19 diagnostic biomarkers using machine learning algorithms, and our study, in combination with existing commercial nucleic acid detection kits, promises to improve COVID-19 detection accuracy.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XS and JZ: conceptualization. XT: methodology and investigation. WY: software and resources. XS, JZ, and WC: validation. WC and QW: formal analysis. DS: data curation and supervision. All authors: writing—original draft preparation and writing—review and editing. QW: visualization. XS: project administration. WC: funding acquisition. Funding acquisition: WC. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by Jiaxing Fight Novel Coronavirus Pneumonia Emergency Technology Attack Special Project in 2020 (No. 2020GZ30001), the Key Discipline of Jiaxing Respiratory Medicine Construction Project (No. 2019-zc-04, 2019-zc-12), General Scientific Research Project of Education Department of Zhejiang Province (No. Y202043729) and Jiaxing Key Laboratory of Precision Treatment for Lung Cancer.

ACKNOWLEDGMENTS

We would like to thank the researchers and study participants for their contributions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.926069/full#supplementary-material

Supplementary Table S1 | Thirty-seven feature genes obtained based on the XGBoost algorithm.

REFERENCES

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China 2019. *N Engl J Med.* (2020) 382:727–33. doi: 10.1056/NEJMoa2001017
2. Yuce M, Filiztekin E, Ozkaya KG. COVID-19 diagnosis-a review of current methods. *Biosens Bioelectron.* (2021) 172:112752. doi: 10.1016/j.bios.2020.112752
3. Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, et al. Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *Lancet Infect Dis.* (2020) 20:678–88. doi: 10.1016/S1473-3099(20)30162-6
4. Salathe M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss Med Wkly.* (2020) 150:w20225. doi: 10.4414/smw.2020.20225
5. Loeffelholz MJ, Tang YW. Laboratory diagnosis of emerging human coronavirus infections - the state of the art. *Emerg Microbes Infect.* (2020) 9:747–56. doi: 10.1080/22221751.2020.1745095
6. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* (2020) 25:45. doi: 10.2807/1560-7917.ES.2020.25.3.2000045
7. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 CASES. *Radiology.* (2020) 296:E32–40. doi: 10.1148/radiol.2020200642

8. Usman M, Gunjan VK, Wajid M, Zubair M, Siddiquee KN. Speech as A Biomarker for COVID-19 detection using machine learning. *Comput Intell Neurosci.* (2022) 2022:6093613. doi: 10.1155/2022/6093613
9. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: Association for Computing Machinery (2016). p. 785–94.
10. Zhang A, Liu CR, Lin GL. P53 pathway activate detection based on machine learning: The modified XGBoost-based method of pan-cancer pathway activity detection in the cancer genome atlas. In: *CCEAI 2021: 5th International Conference on Control Engineering and Artificial Intelligence*. New York, NY: Association for Computing Machinery (2021).
11. Athanasiou M, Sfrintzeri K, Zarkogianni K, Thanopoulou AC, Nikita KS. An explainable XGBoost-based approach towards assessing the risk of cardiovascular disease in patients with Type 2 Diabetes Mellitus 2020. In: *IEEE 20th International Conference on Bioinformatics and Bioengineering*. Cincinnati, OH: IEEE (2020).
12. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
13. Nakamura M, Kajiwara Y, Otsuka A, Kimura H. LVQ-SMOTE - learning vector quantization based synthetic minority over-sampling technique for biomedical data. *BioData Min.* (2013) 6:16. doi: 10.1186/1756-0381-6-16
14. Gui T, Dong X, Li R, Li Y, Wang Z. Identification of hepatocellular carcinoma-related genes with a machine learning and network analysis. *J Comput Biol.* (2015) 22:63–71. doi: 10.1089/cmb.2014.0122
15. Liu H, Setiono R. Incremental feature selection. *Appl Intellig.* (1998) 9:217–30. doi: 10.1023/A:1008363719778
16. Yang F, Wang X, Ma H, Li J. Transformers-sklearn: a toolkit for medical language understanding with transformer-based models. *BMC Med Inform Decis Mak.* (2021) 21:90. doi: 10.1186/s12911-021-01459-0
17. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* (2020) 21:6. doi: 10.1186/s12864-019-6413-7
18. Yang L, Qin Y, Jian C. Screening for core genes related to pathogenesis of Alzheimer's disease. *Front Cell Dev Biol.* (2021) 9:668738. doi: 10.3389/fcell.2021.668738
19. Tal Galili AOC. Jonathan Sidi, Carson Sievert. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics.* (2017) 34:1600–2. doi: 10.1093/bioinformatics/btx657
20. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
21. Feng C, Wang L, Chen X, Zhai Y, Zhu F, Chen H, et al. A novel artificial intelligence-assisted triage tool to aid in the diagnosis of suspected COVID-19 pneumonia cases in fever clinics. *Ann Transl Med.* (2021) 9:201. doi: 10.21037/atm-20-3073
22. Kukar M, Guncar G, Vovko T, Podnar S, Cernelc P, Brvar M, et al. COVID-19 diagnosis by routine blood tests using machine learning. *Sci Rep.* (2021) 11:10738. doi: 10.1038/s41598-021-90265-9
23. Gao X, Liu Y, Zou S, Liu P, Zhao J, Yang C, et al. Genome-wide screening of SARS-CoV-2 infection-related genes based on the blood leukocytes sequencing data set of patients with COVID-19. *J Med Virol.* (2021) 93:5544–54. doi: 10.1002/jmv.27093
24. Park A, Harris LK. Gene expression meta-analysis reveals interferon-induced genes associated with SARS infection in lungs. *Front Immunol.* (2021) 12:694355. doi: 10.3389/fimmu.2021.694355
25. Prasad K, Khatoon F, Rashid S, Ali N, AlAsmari AF, Ahmed MZ, et al. Targeting hub genes and pathways of innate immune response in COVID-19: a network biology perspective. *Int J Biol Macromol.* (2020) 163:1–8. doi: 10.1016/j.ijbiomac.2020.06.228
26. Li J, Liu S, Shi J, Zhu HJ. Activation of tenofovir alafenamide and sofosbuvir in the human lung and its implications in the development of nucleoside/nucleotide prodrugs for treating SARS-CoV-2 pulmonary infection. *Pharmaceutics.* (2021) 13:656. doi: 10.3390/pharmaceutics13101656
27. Vastrad B, Vastrad C, Tengli A. Bioinformatics analyses of significant genes, related pathways, and candidate diagnostic biomarkers and molecular targets in SARS-CoV-2/COVID-19. *Gene Rep.* (2020) 21:100956. doi: 10.1016/j.genrep.2020.100956
28. Lapointe CP, Grosely R, Johnson AG, Wang J, Fernandez IS, Puglisi JD. Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation. *Proc Natl Acad Sci U S A.* (2021) 118:118. doi: 10.1073/pnas.2017715118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Song, Zhu, Tan, Yu, Wang, Shen and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

SYSTEMATIC REVIEW
published: 23 June 2022
doi: 10.3389/fpubh.2022.869238

A Survey on Machine Learning and Internet of Medical Things-Based Approaches for Handling COVID-19: Meta-Analysis

Shahab S. Band[1], Sina Ardabili[2], Atefeh Yarahmadi[1], Bahareh Pahlevanzadeh[3], Adiqa Kausar Kiani[1], Amin Beheshti[4], Hamid Alinejad-Rokny[5,6,7], Iman Dehzangi[8,9], Arthur Chang[10]*, Amir Mosavi[11,12]* and Massoud Moslehpour[13,14]**

[1] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Douliou, Taiwan, [2] Department of Informatics, J. Selye University, Komárom, Slovakia, [3] Department of Design and System Operations, Regional Information Center for Science and Technology (R.I.C.E.S.T.), Shiraz, Iran, [4] Department of Computing, Macquarie University, Sydney, NSW, Australia, [5] BioMedical Machine Learning Lab, The Graduate School of Biomedical Engineering, U.N.S.W. Sydney, Sydney, NSW, Australia, [6] U.N.S.W. Data Science Hub, The University of New South Wales (U.N.S.W. Sydney), Sydney, NSW, Australia, [7] Health Data Analytics Program, AI-enabled Processes (A.I.P.) Research Centre, Macquarie University, Sydney, NSW, Australia, [8] Department of Computer Science, Rutgers University, Camden, NJ, United States, [9] Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, United States, [10] Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology, Douliu, Taiwan, [11] John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary, [12] Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia, [13] Department of Business Administration, College of Management, Asia University, Taichung, Taiwan, [14] Department of Management, California State University, San Bernardino, CA, United States

OPEN ACCESS

Edited by:
Ali Kashif Bashir, Manchester Metropolitan University, United Kingdom

Reviewed by:
Taki Hasan Rafi, Hanyang University, South Korea
Umut Özkaya, Konya Technical University, Turkey
Qasem Abu Al-Haija, Princess Sumaya University for Technology, Jordan
Ali Mohammad Alqudah, University of Manitoba, Canada

***Correspondence:**
Shahab S. Band
shamshirbands@yuntech.edu.tw
Arthur Chang
changart@yuntech.edu.tw
Amir Mosavi
amir.mosavi@kvk.uni-obuda.hu
Massoud Moslehpour
writetodrm@gmail.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *04 February 2022*
Accepted: *20 April 2022*
Published: *23 June 2022*

Citation:
Band SS, Ardabili S, Yarahmadi A, Pahlevanzadeh B, Kiani AK, Beheshti A, Alinejad-Rokny H, Dehzangi I, Chang A, Mosavi A and Moslehpour M (2022) A Survey on Machine Learning and Internet of Medical Things-Based Approaches for Handling COVID-19: Meta-Analysis. Front. Public Health 10:869238. doi: 10.3389/fpubh.2022.869238

Early diagnosis, prioritization, screening, clustering, and tracking of patients with COVID-19, and production of drugs and vaccines are some of the applications that have made it necessary to use a new style of technology to involve, manage, and deal with this epidemic. Strategies backed by artificial intelligence (A.I.) and the Internet of Things (IoT) have been undeniably effective to understand how the virus works and prevent it from spreading. Accordingly, the main aim of this survey is to critically review the ML, IoT, and the integration of IoT and ML-based techniques in the applications related to COVID-19, from the diagnosis of the disease to the prediction of its outbreak. According to the main findings, IoT provided a prompt and efficient approach to tracking the disease spread. On the other hand, most of the studies developed by ML-based techniques aimed at the detection and handling of challenges associated with the COVID-19 pandemic. Among different approaches, Convolutional Neural Network (CNN), Support Vector Machine, Genetic CNN, and pre-trained CNN, followed by ResNet have demonstrated the best performances compared to other methods.

Keywords: machine learning, COVID-19, Internet of Things (IoT), deep learning, big data, information systems, internet of medical things, coronavirus

INTRODUCTION

The outbreak of COVID-19 in Wuhan City, Hubei Province, China, began in December 2019 through the seafood wholesale market (1). Later, on January 30, 2020, the World Health Organization (WHO) declared the prevalence of Covid-19 as an emergency pandemic worldwide (2). Many governments have declared it a dangerous pandemic and imposed full quarantine to prevent the spread of COVID-19. Several countries have reduced their growing infection by

tightening quarantine and forcing people to maintain social distance (3). Even if through complete quarantine, they failed to control the COVID-19 completely. Some countries have joined in the medical development to treat COVID-19. However, to date, there is no specific drug to treat COVID-19. However, few drugs have been suggested as potential research therapies. The proposed drug has been studied under WHO-led clinical trials (4). According to several studies, since COVID-19 is a communicable disease, the WHO has stated that complete quarantine could be the only way to prevent COVID-19 (5).

The COVID-19 outbreak has created many challenges in human life worldwide (6). The most devastating impact, increasing casualties and deaths (around the world), has made it clear the need for social and business restrictions (7). With the expansion of the COVID-19 pandemic, the world community has faced many other problems in various aspects of life, such as economic and social life, psychological wellness, political interactions, cultural activities, educational limitations, religious restrictions, and even sports events (8, 9). Such examples highlight the need for effective and intelligent systems to deal with such crises in the pandemic situation (9). Early diagnosis, prioritization, screening, clustering and tracking of patients, and production of drugs and vaccines are some of the applications that have made it necessary to use a new style of technology to involve, manage, and deal with this epidemic (10). Machine Learning (ML) and Artificial intelligence (AI) algorithms displayed promising ability in prediction and classification (11–22) including disease prediction (23–34), virus genome analysis (24, 35, 36), and medical imaging and Internet of Things (37–40). Strategies backed by artificial intelligence (A.I.) and the Internet of Things (IoT) have been undeniable to understand how the virus works and try to prevent it from spreading (9, 41). These techniques have evolved with the development of computing resources with cloud computing and recent advances in ML. These advances enable researchers to process large amounts of data and extract information. ML-based methods used in processing and modeling data on COVID-19 disease can increase efficiency and speed up results by improving computations. Several researchers have moved toward using ML-based techniques for different applications in the COVID-19 dataset, such as classification using C.T. Images (42), chest C.T. Images (43), and X-ray images (44).

Given the diversity of data, applications, and even the multiplicity of machine learning methods, it is necessary to develop a comprehensive survey study that can consider all the strengths and weaknesses in a standard and systematic study. **Table 1** presents similar survey studies developed in the field for describing their ability to convey their message on the subject reviewed. **Table 1** discusses the study's strengths to find the main research gap.

In one of the early studies, Gou et al. presented a survey to evaluate the ML-based techniques for diagnosing COVID-19 using medical data collection, image preprocessing, feature extraction, and image classification. The study evaluates Transfer, ensemble, unsupervised and semi-supervised learnings, convolutional neural networks, graph neural networks, and explainable deep neural networks. Evaluations focused on the advantages and limitations of the diagnosis techniques (45). Abumalloh et al. presented a state-of-the-art ML-based technique for handling medical image processing in the context of the COVID-19 crisis (46). Khan et al. developed a survey of the applications of A.I. for preventing the COVID-19 pandemic (47). El-Rashidy et al. conducted a review study to describe A.I.'s role in preventing the COVID-19 pandemic using the five applications, including COVID-19 diagnosis, estimation of the COVID-19 outbreak, and patient characteristics, as well as vaccine development (48). Later on, Alballa and Turaiki surveyed the recent articles on ML techniques for COVID-19 diagnosis, mortality rate prediction, and violence risk estimation (49). As can be deduced, many survey studies have been developed. But, the existence of a study that can systematically review and discuss two interrelated areas of the ML and the IoT in the form of an article has been lost from the research literature.

TABLE 1 | The description of the conducted review articles.

Reference	Highlights	Database information	Probable gap
Guo et al. (45)	ML for COVID-19 Diagnosis	NA.	Limited field of the study and lack of proper database information
Abumalloh et al. (46)	ML methods for processing the medical image in the context of the COVID-19 crisis	Eight electronic databases: Elsevier, IEEE, PubMed, Wiley Online Library, Springer, Summon, Google Scholar, and Taylor and Francis	Limited field of the study and proper evaluation
Khan et al. (47)	AI for preventing the COVID-19 pandemic	ScienceDirect, Google Scholar, and preprints from arXiv, medRxiv, and bioRxiv	Subject review interval and evaluation of methods
El-Rashidy et al. (48)	The role of A.I. in preventing the COVID-19 pandemic	Textual data, medical images, and speech data	The subject review interval
Alballa and Al-Turaiki (49)	ML techniques for COVID-19 diagnosis, mortality, and violence risk estimation	PubMed, Scopus, IEEE Xplore, and Google Scholar	Limited subject review interval

The main contribution of this study is to systematically investigate and analyze the role of ML and the Internet of Medical Things (IoMT) to address the challenges associated with diagnosis of the COVID-19 and its outbreak prediction. Here we comprehensively investigate the merits and shortcomings of the ML and IoMT tools proposed for these tasks and present a numerical and statistical analysis.

There is an urgent need to utilize existing technologies to their full potential. Internet of Things (IoT) and ML is regarded as one of the most trending technologies with great potential in fighting against the coronavirus outbreak. The IoT comprises a scarce network in which the IoT devices sense the environment and

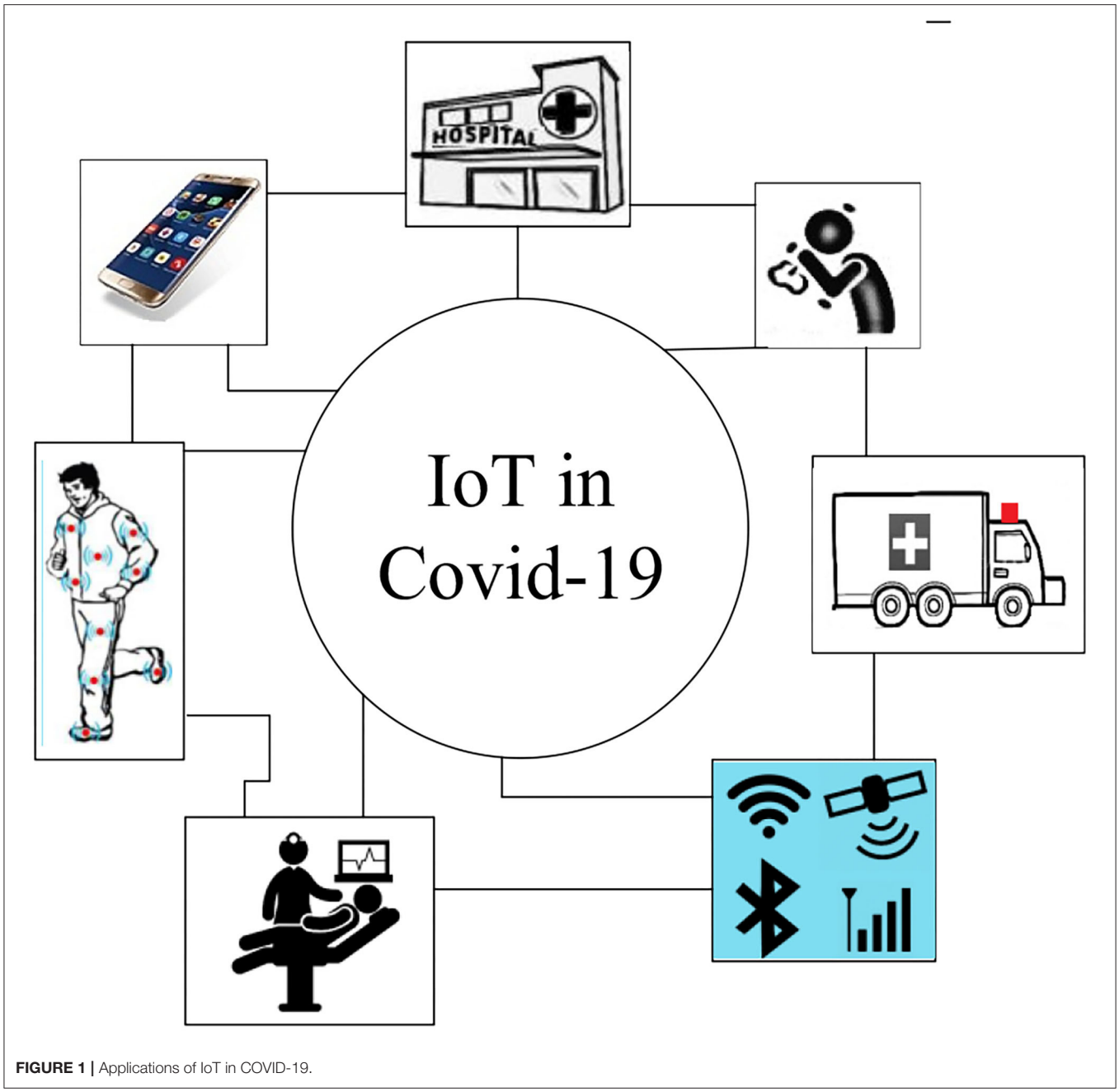


FIGURE 1 | Applications of IoT in COVID-19.

send valuable data on the internet. In this review, we examine the current status of IoT applications and ML related to COVID-19, identify their deployment and operational challenges, and suggest possible opportunities to contain the pandemic further.

The IoT provides the materials needed to help the world minimize the effects of COVID-19. The Internet of Things works with a wide range of applications to ensure compliance with health authorities' safety instructions and precautions. The Internet of Things has a scalable network with the potential to deal with the vast amount of data received from sensors used by several programs to combat COVID-19. In addition, reliable IoT networks reduce critical data delivery times, which can help provide a timely response during the global COVID-19 epidemic. Due to the prevalence of the COVID-19, the role of the Internet of Things was never as needed as it is now.

Artificial intelligence (A.I.) is one of the most important and promising technologies that help revolutionize many fields by creating a revolution. The introduction of machine learning algorithms and artificial intelligence to the Internet of Things has opened new doors in this field. Machine learning provides the opportunity to learn and extract meaningful patterns from data. Because IoT device data is collected in a database, it can easily be

TABLE 2 | The main studies for the application of IoT based techniques for handling COVID-19.

Order	Objective	Methodology/ Proposed Algorithms	Dataset	Keywords	Application	Reference
1	To aim an innovative IoT-based online solution for tracking COVID-19 outbreaks	IoT-based platform to contact and to trace the infection	5G wireless, cloud technologies, and largescale data	I.O.T.: symptom-based device-to-device (D2D) communication	Prediction and monitoring	(66)
2	To compare DL techniques to detect COVID-19	DL-based COVID-19 diagnosis technique in order to model instances for each type and to diagnosis the vulnerabilities	Data from medical IoT devices	IOT: DL algorithm, AE	Diagnostic	(67)
3	To develop an IoT-based DL platform for early detection of COVID-19	Chest X-Ray pictures for training and testing of Regional-based Convolutional Neural Networks (R.C.N.N.) through IoT-based framework	Chest X-Ray images	IoT, COVID-19, Deep learning, Region Proposal Network (RPN)	Diagnostic	(68)
4	To develop a monitoring and detection system according to real-time data from in the presence of the machine learning algorithms	SVM, ANN, Naïve Bayes, K-NN, DT, Decision Stump, 1-R, and 0-R.	Actual COVID-19 patient identifiers include: Fever, Cough, Fatigue, Sore Throat, and Shortness of Breath	Machine learning algorithms, COVID-19	Identification and monitoring	(69)
5	To investigate the IoT for diagnosis of COVID-19 patients using interconnected network	12 IoT based monitoring systems are identified and discussed.	Dataset from databases of Google Scholar, PubMed, S.C.O.P.U.S. and ResearchGate	Internet of things (IoT)	Monitoring	(70)
6	To investigate participants' health conditions and remembering the maintain physical distancing	A lightweight and low-cost IoT node using a smartphone, and fog-based ML for data handling	Vital data from participations	Internet of Things (IoT), smartphone application, Machine Learning (ML), Fuzzy system	Monitoring	(71)
7	To aim a smart edge monitoring system using smart gadgets	To diagnose coronavirus infection using gadgets, deep edge computing and IoT to detect the virus suspected H2H chain	Data from sensors	COVID-19, Edge Computing, IoT	Monitoring	(72)
8	To employ a non-contact I.R. sensor to evaluate for the body temperature	Checking the health condition	Body temperature	IoT, detection system	Detection	(73)
9	To develop a Medical Diagnosis Humanoid to provide a complete diagnostic system for COVID-19	Autonomous navigation, detection, and monitoring system	Data from six different health modules	IoT, A.I., ML, Medical Diagnosis Humanoid	Monitoring and Diagnosis	(74)
10	To develop a low-cost robotic system to diagnosis and help virus affected people	To track hand gestures using radio frequency	Hand gesture	Wireless Robot, Gesture Recognition, IoT	Diagnosis and monitoring system	(75)
11	To contribute IoT and associated sensor technologies to trace, track and mitigate COVID-19 virus by developing hardware sensor	to integrate IoT techniques and provide insight on the expected outcomes	Temperature, Location, Imaging, Pay-point data, and Social media feeds dataset	A.I., IoT, big data, data sharing, cloud computing	Diagnosis and monitoring system	(76)
12	To extract the social relationships between mobile devices by allocating the limited protective resources	To employ dynamic W.U.G. model using social IoT	Pair of real-life datasets	Social Internet of Things; susceptible-exposed-infected-removed; reinforcement learning	Detection	(77)

(Continued)

TABLE 2 | Continued

Order	Objective	Methodology/ Proposed Algorithms	Dataset	Keywords	Application	Reference
13	To develop Internet of Things to prevent the spreading of COVID-19	Investigating an infected person using IoT	NA.	Internet of Things; health care; cloud computing	Detecting and Monitoring	(78)
14	To develop a platform for biometric face detection along with COVID-19 outbreaks	IoT-based Multi-task Cascaded Convolutional Network	Face image dataset	Detection, cascaded CNN, cloud computing, IoT, edge computing,	Detection and recognition	(79)
15	To introduce a high resolution A.Q. monitoring system	A preliminary validation of the Air Heritage pervasive Air Quality monitoring concept	Air quality dataset	Smart Air Quality monitors, IoT, Artificial Intelligence, COVID-19,	Monitoring	(80)
16	To develop an IoMT architecture with respect to combat COVID-19.	IoMT platform, emerging IoMT applications, to apply within the medical environment	N.A.	COVID-19, IoMT application, security	Detection	(38)
17	To test information technology for handling the COVID-19 pandemic	A.I., block-chain, Big Data and robots, for optimally handling pandemics	Google Scholar database and Proquest	COVID-19, information technology, A.I., big data, indonesia	Detection and monitoring pandemic	(81)

used to predict the prevalence and effects of the coronavirus and how to reduce it. Data of patients with COVID-19 help predict the future behavior of the virus and regional comparison of its effects. In addition, it also helps with the possible adaptation of COVID-19 symptoms to an effective and rapid A.I. treatment.

The patient's medical record and the results obtained help to predict better treatment choices based on artificial intelligence and machine learning (ML) algorithms and lead to rapid recovery and patient monitoring. Artificial intelligence-based emergency traffic control paves the way for ambulances and other emergency service providers. BlueDot was one of the first artificial intelligence companies to predict the outbreak of the Corona virus and identify its global threat. They provided information on the mobility pattern of the virus and its potential for spread. Other A.I. companies also joined hands to work with COVID-19, including Deargen, Insilico Medicine, and S.R.I. Biosciences and Iktos, Benevolent AI, DeepMind, Nanox, Baidu, Alibaba, and EndoAngel Medical Technology Co.

Here we conclude that there is a gap in how to address the strengths and weaknesses of machine learning and IoT methods that need to be addressed. In the meantime, to close this gap, we will need to classify, determine the pros and cons, challenges and limitations, and outline ways to deal effectively with COVID-19. In line with this basic need to have a deeper insight into the applications and effects of machine learning and the Internet of Things on the COVID-19 Pandemic, we presented research to be able to study these methods in different ways and in a practical way.

Accordingly, the main purpose of this review article is to examine the ML, IoT, and the integration of IoT and ML-based techniques in the applications related to COVID-19 from the diagnosis of the disease to the prediction of its outbreak.

The study has three main sections:

- A section for describing the studies developed by IoT and IoT-ML based techniques in COVID-19 applications.
- A section for presenting the role of ML-based techniques in COVID-19 applications.
- A section for presenting the main findings, challenges, and future perspectives.

METHODOLOGY

Dataset Preparation Method

A systematic review may provide technical and practical literature for a specific topic (50). A systematic review requires a proper collection of papers on the subject. Preparing a dataset is one of the main steps in determining review work quality (50). According to the P.R.I.S.M.A. guidelines, the present study has collected the most relevant studies from W.O.S. and Scopus libraries (51). There are four steps for preparing the database using the P.R.I.S.M.A. guidelines, including (52): (1) identification, (2) screening, (3) eligibility, and (4) inclusion. In the identification step, we employed the frequently used keywords, including COVID-19, pandemic, diagnosis, detection, Prediction, Monitoring, Classification, Identification, IoT, and Machine learning to search within the article title abstract and keywords. In the first step, about 109 articles have been identified and selected from the Thomson Reuters Web-of-Science (WoS) and Elsevier Scopus. In the Screening step, the duplicate articles have been removed. Twenty-five articles (about 23%) have been removed from the dataset. Thirty nine articles (about 36%) have been eliminated due to the lack of details on the methods, datasets, full text, etc. A total of 45 cases (about 41%) have been included in the screening step. In the eligibility step, the authors team surveyed the full text of the papers, and marked the relevant articles during monitoring eligibility. In this phase,

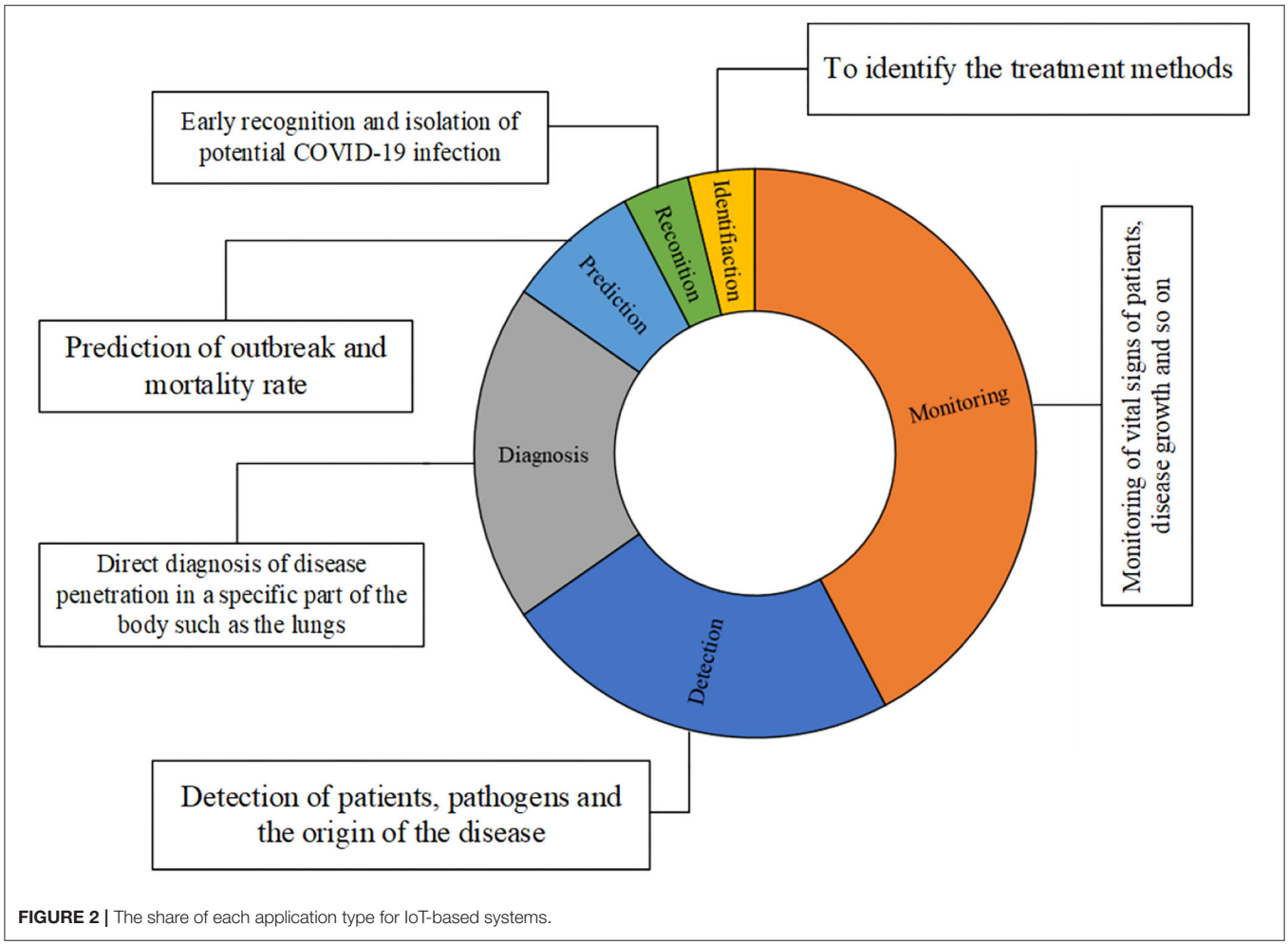


FIGURE 2 | The share of each application type for IoT-based systems.

23 cases (about 22%) have been selected for investigating the evaluation criteria and including in the dataset.

IoT for COVID-19

IoT is an interconnected set of computing tools from simple to complex that can be used in conjunction with mechanical or digital machines in the presence of humans, animals, or objects. IoT technology can easily transfer data from the source to the destination through the network without the operator's presence. This technology can be considered a special tool in human-human interaction or human-computer interaction (53, 54). An IoT platform includes the minimum equipment required, such as smart devices equipped with the web (55). These systems consist of processors, sensors, and communication hardware to collect, send, control, manage, and convert data into accessible data (55, 56). These systems connect to an IoT port used to send data to the cloud so that data can be analyzed and shared (57). These devices can operate by connecting to other related systems based on their information (58). These tools perform many of their tasks without human intervention.

Today, IoT technology in health and treatment is growing rapidly (59). The main applications of IoT in the process of treatment and intelligent health can include identification, digitization of medical information, patient transfer to the hospital, use of vital signs sensors, use of smartphones in communication, and digitization of medical processes (60, 61). Furthermore, IoT has become more popular and important due to the COVID-19 pandemic (62). Because this virus is highly contagious and has a high risk to human health, and has caused many problems for the medical staff, using non-contact methods to diagnose as soon as possible, control patients, monitor the condition of patients with acute illness, as well as maintain social distance, can be an important factor in breaking off part of the virus infection cycle (63, 64). In non-contact methods, the IoT is a leader and can solve many problems in this field (65). **Figure 1** presents the main applications of the IoT in COVID-19 era.

Table 2 presents the highlighted studies for applying IoT-based techniques to tackle COVID-19. This table represents the studies based on the main four columns. First, the objective briefly describes the main objective of each study.

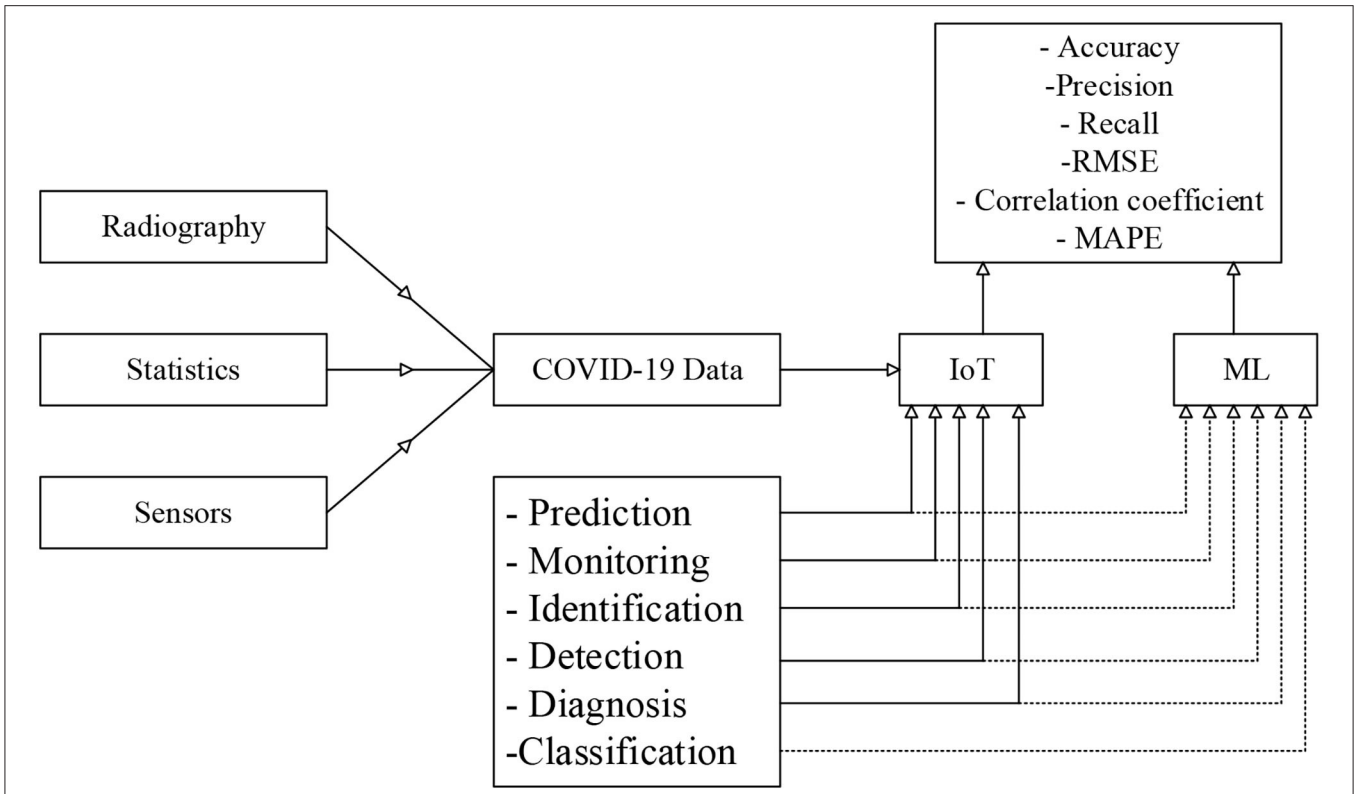


FIGURE 3 | Main contributions of the current study.

TABLE 3 | The main contribution of the study for the application of IoT based techniques.

Methodology	Prediction	Monitoring	Detection	Identification	Diagnostic
IoT	☑	☑	☑		
IoT-DNN					☑
IoT-RCNN					☑
IoT-SVM		☑		☑	
IoT-ANN		☑		☑	
IoT-Naïve Bayes		☑		☑	
IoT-K-NN		☑		☑	
IoT-DT		☑		☑	
IoT-Fog based		☑			
IoT-Deep edge computing		☑			
Wireless sensors		☑			
IoT based S.E.I.R.			☑		
IoT-IT		☑	☑		

Methodology/proposed algorithm presents the main algorithm and procedure employed by each study. Keyword indicates the main points and concentration of the study and finally, the application section presents the field of the application of each method.

Figure 2 presents the contribution of different applications which are performed by IoT techniques to track COVID-19 related fields.

Figure 2 is generated based on **Table 2** to present the main applications and their portions by studies for handling IoT in

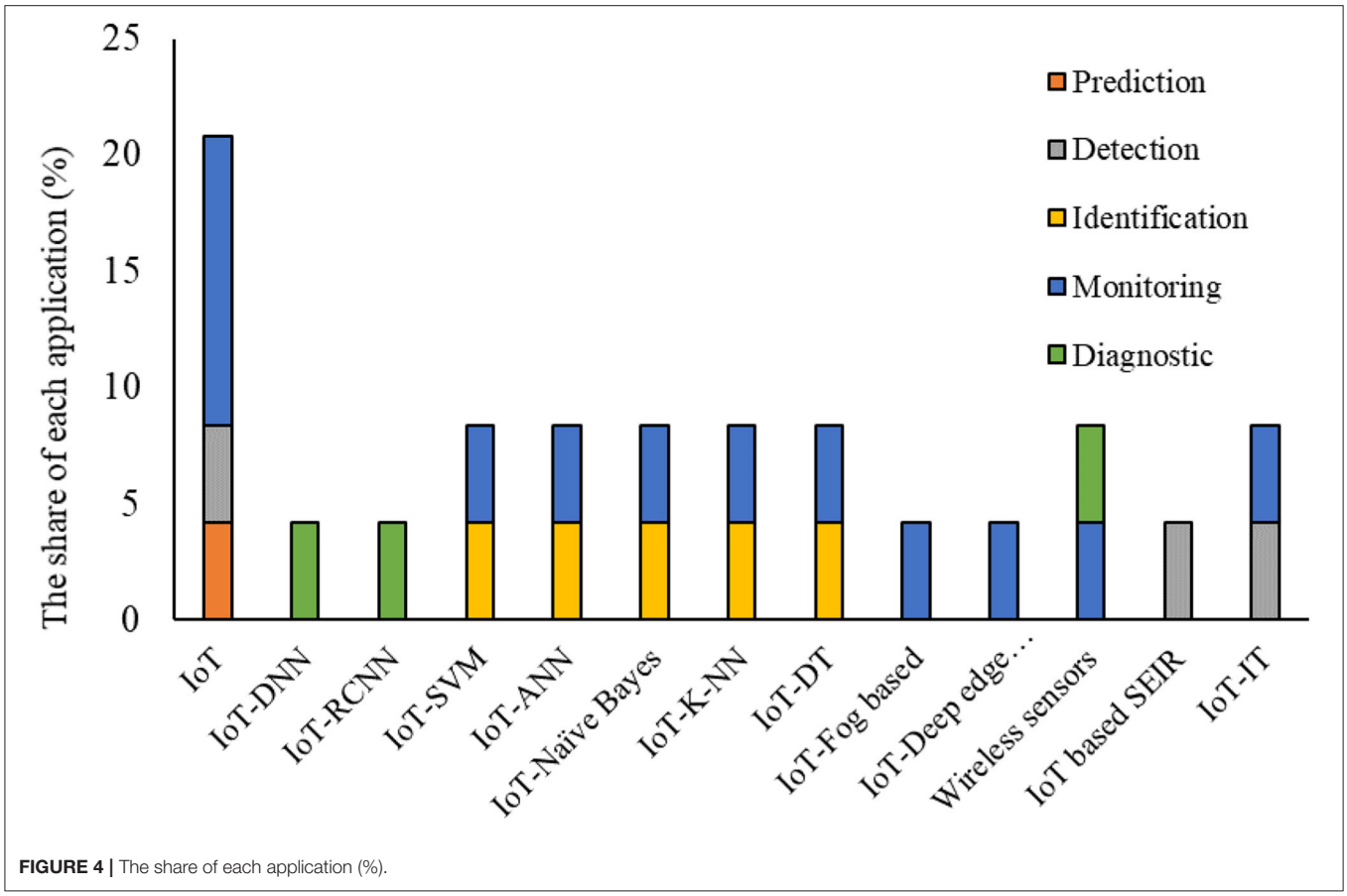


FIGURE 4 | The share of each application (%).

the COVID-19 pandemic. Monitoring, detection, and diagnosis are the main application of IoT-based techniques in tackling the COVID-19 pandemic. Monitoring can be performed in different ways. Accordingly, Roy et al. employed IoT as a real-time solution for monitoring COVID-19 outbreaks (66). Also, Otoom et al. employed IoT to provide monitoring and detection data using a real-time system to feed to the machine learning algorithms for further applications or handling (69). Singh et al. and Vedaei et al. used IoT as a tool for monitoring COVID-19 patients and their health condition in cooperating with an interconnected network (70, 71). Ashraf et al. proposed a smart edge surveillance system to monitor wearable smart gadgets which are operated according to IoT-based technology (72). Karmore et al. developed a Medical Diagnosis Humanoid to provide a complete diagnostic system for COVID-19 using IoT-based technology (74). De Vito et al. presented the outputs of a high-resolution A.Q. monitoring system based on an IoT-based technique (80).

Baskaran et al. used a non-contact infrared sensor to examine the body temperature to detect the patients with COVID-19 (73). Wang et al. exploited the social relationships in the platform of Social IoT to solve controlling issues of the COVID-19 epidemic by sharing the limited protective resources (77). Kumar et al. investigated an IoT based platform to prevent the spreading of COVID-19 (78). Kolhar et al. developed a platform of a decentralized IoT-based biometric based on a face detection platform for handling COVID-19 outbreaks (79). Aman et al. developed an architecture of IoT based framework for medical applications with respect to combat COVID-19 (38). Manalu et al. investigated the information technology to respond COVID-19 pandemic trend in accordance with the IoT technology (81).

Figure 3 presents the main contribution of these papers. According to the reviewed studies, the COVID-19 dataset can be imported from three main sources, including Radiography, statistics of health centers, and Sensors for prediction, monitoring, identification, detection, diagnosis, and classification purposes. The output of the techniques needs to be evaluated to confirm the approach performance and accuracy values. The frequently used parameters for performance analysis include Accuracy, Precision, Recall, R.M.S.E., Correlation coefficient and mean absolute percentage error. This can be considered a brief explanation of the main contribution of the present study. This study successfully presents the advantages and disadvantages of each technique for a specific task in handling the COVID-19 dataset and proposes the future perspectives. Also, this study can detect the main challenges and limitations.

There is a need to categorize the main applications of IoT and the relevance technique following COVID-19. **Table 3** presents the study's main contributions to the application of IoT and integrated IoT-ML-based techniques. Based on **Table 3**, the exact application of each of the methods used can be extracted. It is

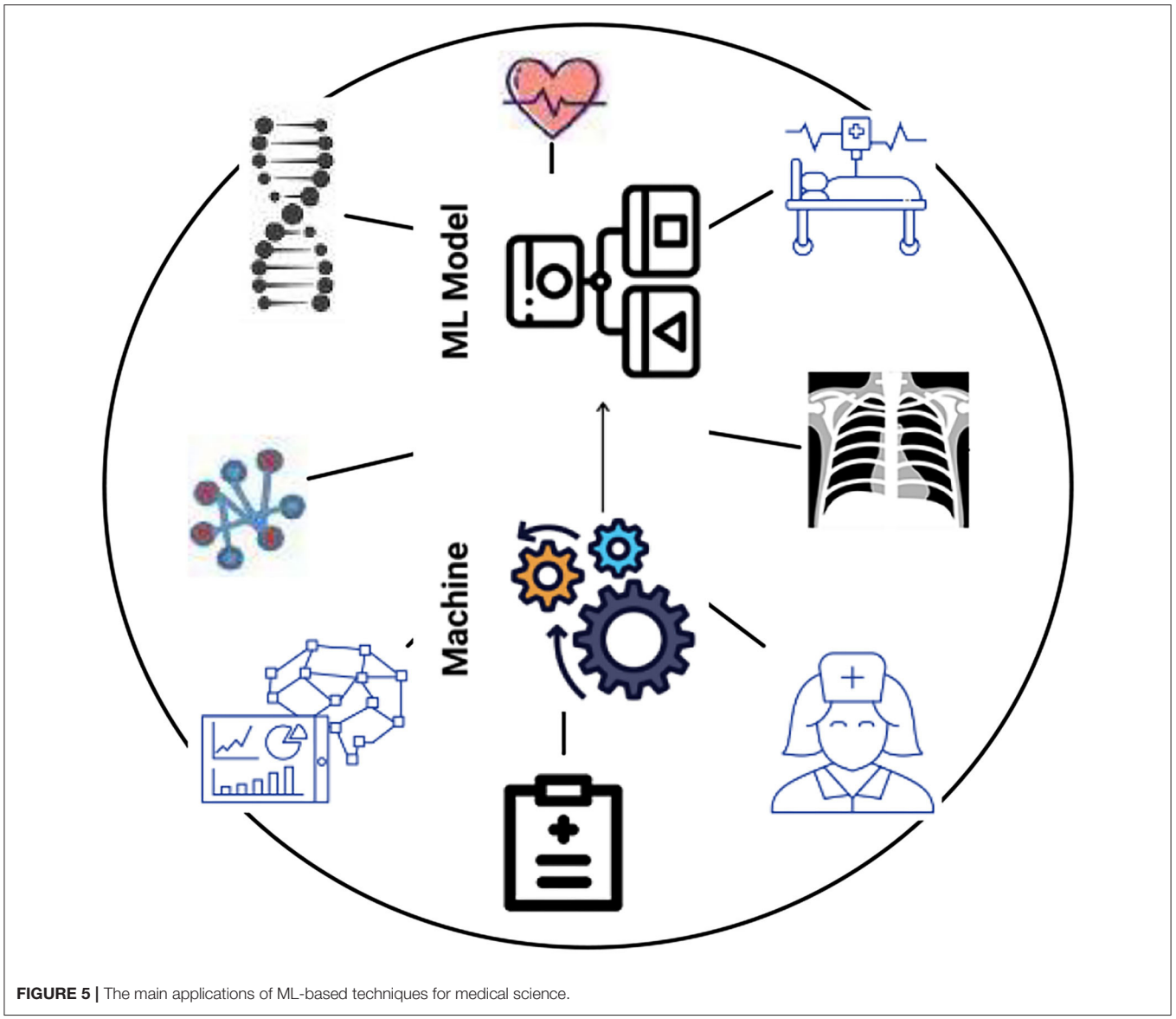


FIGURE 5 | The main applications of ML-based techniques for medical science.

also possible to find out which methodology is still available for which application can be considered a research opportunity for the future. Also, by carefully examining the different reasons for the tendency of each method to the fields shown in independent research, which can be considered necessary research and planning opportunities for policymakers in this field.

As shown in **Table 3**, IoT-based technology requires ML-based techniques to complete the task. **Figure 4** presents the share of each methodology in the applications by percentage.

As shown in **Figure 4**, IoT has been used more than other applications to monitor and detect COVID-19 cases. However, it has been less popular in the identification at the same time.

ML Techniques for Pandemic Prediction of COVID-19

Utilizing the ML platform led to reducing the adverse effects of the disease and accelerating the healing process (62). The combination of A.I. and ML has led to advances in treatment, medication, screening, prognosis, contact tracking, and the drug/vaccine development process and reduced human intervention in medical performance (82). ML is also used as a tool for managing virtual queues to prevent crowds in physical waiting rooms or long queues. In addition, it is used to predict waiting times and implement calls in a privacy manner in conjunction with the cell phone platform (83).

The ML method is widely used in data analysis by intelligently producing an analytical model. This method is a subset of artificial intelligence that analyzes data and produces a model for estimating, categorizing, optimizing, predicting, identifying problems, and decision-making (84, 85).

New computing technologies have made the problems assessed by ML-based techniques today a little different from the way they are analyzed based on past technologies (86). These techniques began to evolve from pattern recognition to

TABLE 4 | ML-based techniques for COVID-19.

Order	Aim	Method	Dataset	Key words	Application	Reference
1	To develop a mask face detection model	Deep transferring learning (ResNet50) as classifier and SVM to be compared with ensemble method	Image-based dataset	Deep transferring learning, SVM, and ensemble	Detection	(92)
2	To employ ML based platform as a healthcare application to proper decision making for COVID-19 detection	Integration of random forest, Gaussian nave bias and Generative adversarial network	Real-time processing of users' health data	Artificial intelligence, Cloud/fog computing, IoT	Detection	(93)
3	To propose an A.I. based technique integrated by C.T. scan and chest x-ray images to identify, and predict the positive infected patients	Pre-trained CNN	Chest X-ray and C.T. scan images	COVID-19, DT, X-ray images, AI	Identification and diagnosis	(94)
4	To employ a novel CNN architecture for classifying COVID-19 from chest X-rays.	CNN architecture	Chest X-ray	DL, CNN, mine data patterns	Classify and identification	(95)
5	To develop an AI based methods for fast diagnosis of COVID-19 cases	ResNet-101 in comparison with Radiology data	Chest X-ray radiography	AI, CNN, ResNet-101	Diagnosis	(96)
6	To detect COVID-19 promptly using CNN	CNN technique	Chest X-ray images	DL, CNN, Squeeze Net	Detection	(97)
7	To develop and test a new computer-aided diagnosis (CAD) to investigate COVID-19	CNN	Multi-center chest C.T. dataset	CNN, DL, CAD	Diagnosis	(40)
8	To propose an intelligence computer-aided model to support daily clinical applications	Convolution neural network (CNN) with SVM classifier architecture on chest X-ray	Chest X-ray	Medical decision support system; Deep learning	Detection	(98)
9	To develop an AI-based model for proper screening and monitoring of COVID-19	AD3D-MIL	Chest X-ray images	Screening, CAD, DL, ML	Monitoring	(99)
10	To present a CNN based technique for early COVID-19 diagnosis from chest X-ray	CNN	Chest X-ray	A.I., CNN, DL	Diagnosis	(39)
11	To investigate a medical decision support system by CNN	CNN	Chest X-ray images	Decision support; CNN; DL; ML	Diagnosis	(100)
12	To propose an intelligent methodology to diagnosis the COVID-19 cases	The multi-criteria decision-making (M.C.D.M.) using T.O.P.S.I.S. in the presence of SVM based classifier	Chest X-ray Dataset	COVID-19 diagnostic, machine learning, benchmarking; TOPSIS,	Diagnosis	(101)
13	To study the utility of A.I. in a prompt and accurate diagnosis of COVID-19 in the presence of chest X-ray images	Pre-trained CNN	Chest X-ray images	AI; COVID-19; machine learning, Convolutional Neural Networks	Diagnosis	(102)
14	ML-based classification approach for handling COVID-19	Extreme gradient boosting (XGBoost) model	Eight pathogenic species	Dinucleotide frequencies, feature representations, genomic signatures, human pathogens, ML, extreme gradient boosting	Classification	(103)

(Continued)

TABLE 4 | Continued

Order	Aim	Method	Dataset	Key words	Application	Reference
15	ML-based classification algorithm for handling infectious diseases, such as COVID-19	KNN, SVM, D.T. and L.R.	Wi-Fi signals data	Machine learning, classification, COVID-19,	Classification	(104)
16	To detect the COVID-19 cases using RNN technique	L.S.T.M. architecture of R.N.N. method for detection based on Cough sound, Breathing sound and voices	Speech and sound analysis dataset	AI, DL, RNN	Detection	(105)
17	To present a fuzzy rule basing system to predict COVID-19 daily cases	Fuzzy rule based	Daily cases data from the Turkish republic health ministry	COVID-19, A.I., fuzzy rule based inference	Prediction	(106)
18	To present a multi-scale discriminative segmentation network to detect COVID-19	MSD-Net	COVID-19 CT segmentation dataset	COVID-19, CT, DL	Diagnosis	(107)
19	To develop a hybrid A.I. technique for the prediction of COVID-19	Integrated natural language processing module and the L.S.T.M.	The epidemic data of several typical provinces and cities in China	COVID-19, prediction, epidemic model, hybrid A.I.,	Prediction, detection	(108)
20	To present a solution for identifying pneumonia using C.X.R. images	GCNN	CXR images	G.C.N.N., Computed Tomography, Chest X-Ray, A.I.	Classification	(109)
21	To examine the emotions expressed by people using social media to track and diagnosis sentiment behind COVID-19	LR, Multinomial Naïve Bayes, DT, RF, SVM and XGBoost classifiers	Fetch data from social media platform	Twitter; emotions; sentiment analysis; pandemic; domain-specific; COVID-19; ML; dataset	Detection	(110)
22	To propose an ML-based approach for the forecasting of COVID-19 cases	M.L.P. and A.N.F.I.S.	Outbreak dataset from WHO	ML, COVID-19 cases, prediction, detection	Detection	(91)
23	To develop hybrid ML-based technique for the globally prediction of COVID-19 cases	Multilayered perceptron integrated by gray wolf optimizer	Outbreak dataset from WHO	Machine learning, COVID-19 cases, prediction, detection	Detection	(90)

a comprehensive theory of the ability of computers to perform specific tasks without the need for special planning (87, 88).

In the field of medicine and treatment, ML is known as one of the most practical tools for analyzing medical data, identifying, predicting, and even treating different situations. With the advancement of medical science in today's world and the production of large volumes of medical data, there is an urgent need to analyze this data (89). **Figure 5** presents the main applications of ML-based techniques for medical science to tackle the COVID-19 pandemic. Identifying the prevalence, effective parameters in the eradication of the virus, identifying patients in the early stages, patients' pattern behaviors, and predicting outbreak and mortality rates can be considered practical and effective areas of ML-based techniques (90, 91).

Table 4 presents the highlighted studies for the application of ML-based techniques for handling COVID-19. Similar to **Tables 3**, **4** discuss them in four columns. The objective column briefly describes the main objective of each study. Methodology/proposed algorithm presents the main algorithm and procedure employed by each study. Keyword indicates the main points and concentration of the study and finally, application section presents the field of the application of each method.

According to **Table 4**, ML-based techniques are employed for detection, identification, monitoring, diagnosis, prediction, and classification purposes in the presence of the COVID-19 dataset. **Figure 6** presents the summary of each application separately. Singh and Kaur employed an ML-based platform using hybrid random forest, Gaussian Naïve Bayes, and Generative adversarial network as a healthcare application to detect COVID-19 cases (93). Vinod et al. developed a pre-trained CNN method as an ML-based technique integrated using C.T. scan and chest x-ray images to identify, detect, and predict the positive infected patients (94). Ardakani et al. developed an ML-based technique (ResNet) for fast diagnosis of COVID-19 cases compared to radiology data (96). Polsinelli et al. developed a study to detect COVID-19 promptly using CNN as a frequently used DL-based architecture (97). Nour et al. proposed an intelligence computer-aided model based on CNN with SVM classifier architecture on chest X-rays to support daily clinical applications (98). Chowdhury et al. investigated the utility of A.I. in the rapid and accurate detection of COVID-19 in the presence of chest X-ray

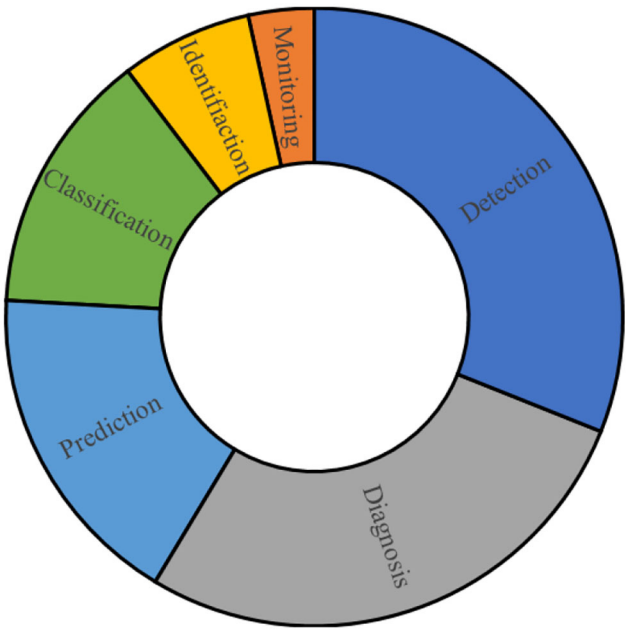


FIGURE 6 | The share of each application type for ML-based systems.

images (102). Sethi et al. employed Logistic Regression (L.R.), Multinomial Naïve Bayes, Decision Tree (D.T.), Random Forest (R.F.), SVM, and XGBoost classifiers to analyze the emotions expressed by people using social media to monitor and detect sentiment behind COVID-19 (110). Ardabili et al. developed ML-based techniques for the prediction of COVID-19 outbreaks (91). In another study, Ardabili et al. also employed a hybrid ML-based technique (Multilayered perceptron integrated by gray wolf optimizer) for the global prediction of COVID-19 cases (90). In addition, Loey et al. employed the DL-based ResNet method in the presence of an SVM-based classifier to detect a masked face (92).

According to **Figure 6**, detection, diagnosis, and prediction can be considered as the main categories of the application of ML-based methods in COVID-19. In general, one of the main sections of analyzing IoT-based and ML-Based techniques applied for a specific field is their evaluation in terms of accuracy, error, or in other word performance of the model. **Table 5** presents the evaluation criteria employed for each model.

According to **Table 5**, accuracy, followed by the recall and precision parameters has owned the highest portion of the evaluation criteria employed for analyzing COVID-19 based dataset using IoT and ML-based techniques. In the following, **Table 6** is generated from **Table 4** for indicating the share of each ML-based technique for each application and their main contributions. According to **Table 6**, ResNet as an architecture of deep learning methods followed by CNN, XGBoost, SVM, D.T., and L.R. has been used more often to tackle work with COVID-19 related data.

Figure 7 presents the share of different ML methods for different tasks to tackle the COVID-19 pandemic. As is clearly indicated in this figure, ResNet, followed by CNN, is the most common application of ML in this field. This can be due to the model's nature for handling different applications like monitoring, detection, identification, classification, and diagnosis. In comparison, other methods can do a limited number of applications.

Evaluation Criteria

Models developed using ML and IoT-ML require an evaluation step for recognizing their performance and accuracy values. According to the studies reviewed, the most effective and frequently used evaluation criteria are including Accuracy, Recall, Precision, Root mean square error (R.M.S.E.), Correlation coefficient and Mean absolute percentage error (M.A.P.E.). These criteria compare the models' output and actual values and provide a comparison score (90, 91). In the present study, we employed the criteria values reported by each study for evaluating and comparing the models. **Table 7** presents the main criteria for evaluation.

Main Findings and Evaluations

This section presents the main findings of IoT based techniques (**Table 8**) and ML-based techniques (**Table 9**). Each table includes two main columns called findings and pros. and cons.

According to **Table 8**, most of these studies lack numerical analysis for the method's performance. One of the main reasons can be the nature of the IoT technique, which goes through a practical process and shows its performance in practical applications and does not need to provide numerical statistics. In all these applications, IoT could successfully cope with the task. IoT provided a fast and efficient approach to tracking the disease spread (66). On the other hand, it can be employed as a real-time framework to minimize the impact of communicable diseases through the early detection of cases (67). In the study by Singh et al. IoT technology successfully increased patient satisfaction and reduced the readmission rate in the hospital (70). However, there is a need to integrate IoT platforms with ML-based techniques for detection purposes. In the study by Rahman et al. DL applications with IoT platforms provided promising findings to detect A.E. attacks. However, there is a need for further research, attention, and implementation of appropriate defense mechanisms, safeguards, and controls (67). Kolhar et al. employed Multi-task Cascaded Convolutional Network architecture (M.C.C.N.N.), and findings claimed that the efficiently integrated by Raspberry Pi increased the robustness of detection and recognition (79).

According to the findings given in **Table 9**, the most share of studies developed by ML-based techniques for handling COVID-19 based datasets provided performance criteria. The most share of the performance criteria, according to **Figure 8** is related to the accuracy factor. Accuracy factor is a general and normalized factor. Therefore, it can be employed for comparing the ML-based methods with different datasets. **Figure 9** presents the accuracy values for each model for comparing their performance in handling the COVID-19 dataset.

Figure 9 indicates CNN with SVM classifier, Genetic CNN, and pre-trained CNN followed by ResNet, provided highest

TABLE 5 | The main evaluation criteria for analyzing the performance of models.

	Accuracy	Recall	Precision	AUC	Sensitivity	specificity	Determination coefficient	RMSE	MAPE	MAE	F1-score
Deep transferring learning (ResNet50)	☑										
RF-NB-GAN		☑	☑								
CNN	☑	☑	☑	☑					☑		☑
ResNet-101	☑				☑	☑					
AD3D-MIL	☑			☑							
T.O.P.S.I.S.							☑				
XGBoost	☑										
kNN	☑										
SVM	☑										
D.T.	☑										
L.R.	☑										
L.S.T.M.	☑			☑							☑
Fuzzy							☑	☑		☑	
MSDN					☑	☑					
Naïve Bayes	☑		☑		☑						
R.F.	☑										
MLP											
ANFIS							☑	☑			
MLP-GWO							☑		☑		
IoT (Medical based)	☑			☑							
Fog-based	☑										
Deep edge computing	☑										
Wireless sensors	☑										
IoT based S.E.I.R.	☑										
I.T.	☑										

TABLE 6 | The main contribution of ML-based techniques in COVID-19 applications.

	Prediction	Monitoring	Diagnosis	Identification	Detection	Classification
ResNet50	☑	☑	☑	☑	☑	☑
RF-Naïve bayes-GDN					☑	
CNN			☑	☑	☑	
ResNet-101			☑			
AD3D-MIL		☑				
T.O.P.S.I.S.			☑			
XGBoost					☑	☑
kNN						☑
SVM					☑	☑
D.T.					☑	☑
L.R.					☑	☑
L.S.T.M.	☑				☑	
Fuzzy	☑					
Multi-scale discriminative network			☑			
G.C.N.N.						☑
Naïve Bayes				☑		
R.F.					☑	
M.L.P.	☑					
A.N.F.I.S.	☑					
MLP-GWO	☑					

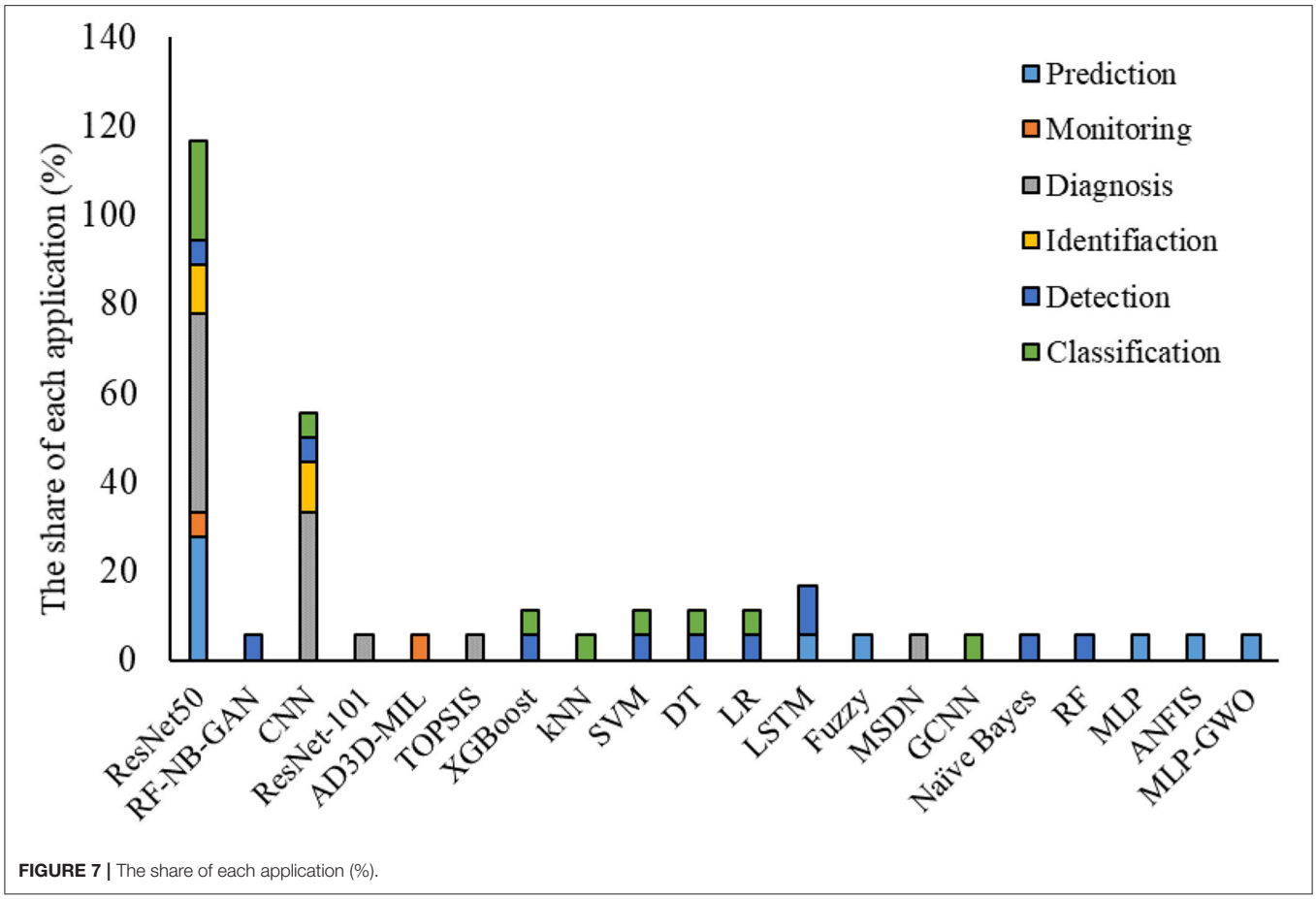


FIGURE 7 | The share of each application (%).

TABLE 7 | The main evaluation metrics.

$Accuracy = \frac{True_p+True_n}{True_p+True_n+False_p+False_n}$	Where $True_p$ denotes the true positives, $True_n$ the true negatives, $False_p$ the false positive, and $False_n$ the false negatives.	(1)
$Recall = \frac{True_p}{True_p+False_n}$	Where $True_p$ denotes the true positives and $False_n$ the false negatives.	(2)
$Precision = \frac{True_p}{True_p+False_p}$	Where $True_p$ denotes the true positives and $False_p$ the false positives.	(3)
$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}$	Where N denotes the total number of samples, x_i the actual samples, and $\hat{x}_i$ the predicted samples.	(4)
$Correlation\ Coefficient = \frac{Cov(x,\hat{x})}{\sigma_x\sigma_{\hat{x}}}$	Where x refers to actual samples, $\hat{x}$ to predicted samples, $Cov(x, \hat{x})$ to the covariance between x and $\hat{x}$, and σ to the standard deviation (calculated for both x and $\hat{x}$)	(5)
$MAPE = \frac{100}{N}\sum_{i=1}^{N}\lvert\frac{x_i-\hat{x}_i}{x_i}\rvert$	Where N denotes the total number of samples, x_i the actual samples, and $\hat{x}_i$ the predicted samples.	(6)

accuracy values. On the other hand, the lowest accuracy was related to single CNN followed by XGboost and K.N.N. techniques.

Challenges and Limitations

Nowadays, when the world is struggling with COVID-19 disease, every innovation and technology is used to fight this disease. Like many other areas, healthcare requires the support of new technologies such as IoT, and ML. Exploring the disease-related dataset, data preparation, prevention, and control of infectious diseases has become one of the main purposes of A.I. IoT and ML have a vital personality in better understanding, dealing with the COVID-19 crisis, and discovering the COVID-19 vaccine. ML-based technology allows computers to predict the pattern and speed of disease transmission with their intelligence and by mimicking large amounts of data. A.I. uses information from people with coronary heart disease, and improved and dead people as tracking data.

To combat the spread of the corona virus, IoT-based methods of communicating with patients provide transparency and a better understanding of how the virus is spread and strengthen

TABLE 8 | The main findings of the study for the application of IoT-based techniques.

Order	Findings	Pros. and Cons.	Reference
1	The proposed solution can identify and track the infected individual and successfully tracks all people who are in the area of disease spread	This framework integrates symptom information as a rapid and efficient approach, thus tracking the prevalence of the disease	(66)
2	DL applications are vulnerable to coronavirus attacks	The method is very vulnerable and requires further studies	(67)
3	The model provides an accuracy of 98% for detection	Combining DL and the IoT makes it easier for radiologists to control the spread of the virus	(68)
4	According to results, all the techniques, except the Decision Stump, OneR, and ZeroR provided accuracies values more than 90%	The proposed platform reduced the communicable diseases using early detection of cases and provided tracking the recovered cases, and a better understanding of the infections	(69)
5	IoT reduces clinical cost and optimizes treatment outcome of the patients	The platform improves patient satisfaction and decreases readmission rate in the hospital	(70)
6	The system can assist tracking the daily activities and decrease the risk of exposure to the COVID-19	The app announces the user to keep a physical distance of 2 m. Also, a Fuzzy-based technique evaluates the environmental risk and user health to estimate the risk of real time spreading. This platform can successfully reduce the coronavirus spread	(71)
7	The platform detects and tracks the infected person	The platform tracks COVID-19 and improves infected person and keeps the dataset for further analysis	(72)
8	The provided package enhances the testing process for increasing the efficiency of the system	This approach will increase the maximum collaboration from the employees	(73)
9	This platform is a cost-effective, safety-critical mobile robotic technology and successfully copes with diagnosis task Also the multiple diagnostic devices increases the detection accuracies	The system effectively provides a complete diagnosis and figuring out COVID-19 patients also contains multiple diagnostic devices, without any need for human interferences	(74)
10	The robot technology protect virus affected persons. The system is also recognizing the patient's Gesture and tracking the instructions	The robot collects data from patient performs tasks without image processing system	(75)
11	IoT-based technology prevent the global pandemic	Improves the control and tracking of a fast-spreading virus such as coronavirus	(76)
12	The proposed methodology is sustainable for disease tracking by an early identification of cases	This technique can successfully handles both governments and other decision-making authorities	(77)
13	This system improves the decision-making procedure	The system is connected through cloud computing and effectively supports the real-time data	(78)
14	Edge computing improved the findings on the decentralized load of face recognition	The platform enhances the robustness of detection and diagnosis	(79)
15	The proposed system could successfully cope with the task	IoT equipped ML can successfully save, and visualize monitoring the volunteers	(80)
16	This study suggests that integrated and hybrid techniques will follow up the near future, using simulation, and forecasting purposes	A higher degree of safety and privacy for humanity	(38)
17	The platform employed for the study have an effective role in the success of pandemic handling	The platform increases accessibility to the proper dataset	(81)

the treatment and research process. ML is one of the new technologies for tracking the spread of the virus and finding effective parameters in it. The ML method can successfully identify high-risk patients and predict the necessary measures to deal with possible infections to reduce the point of the effect of the disease. In addition, ML-based methods can estimate the risk of patient mortality through previous analysis. This technique improves patients' planning, treatment, and reduction and is a complementary medical tool that works with data and evidence. On the other hand, this technology improves decision-making and reduces the cost of treatment and diagnosis. At the same time, in medical imaging, ML tools help to recognize the patterns in the images and strengthen the ability of radiologists to diagnose the possibility of disease and early diagnosis of the disease.

One of the main limitations of IoT, and ML-based techniques for applications in COVID-19 is the lack of a complete dataset. This can be due to the unique development of models by limited data for a specific application within the same data field. The purpose of using IoT, A.I., or ML-based techniques is to solve a specific problem in the real world with a real application that requires the use of special hardware and equipment. There are limitations in the cost and availability of developing and equipping communication hardware in therapeutic, diagnostic, estimation, and forecasting applications for IoT technology or ML-based techniques. Also, there are limited best practices available for IoT developers. The lack of IoT edge authentication and licensing standards has led to restrictions on the application and enactment of laws, regulations, and policies in the use of this technology, and this has led to the absence of IoT-based incident response activities as the best methods. All of these limitations mean that there is still no focus on identifying ways to gain situational awareness of the security of IoT assets in a medical complex.

TABLE 9 | The main findings of the study for the application of ML-based techniques.

Order	Results	Pros. and cons.	Reference
1	The SVM classifier in the presence of R.M.F.D., S.M.F.D. and L.F.W. dataset achieved 99.64, 99.49 and 100% testing accuracy values.	The proposed model provided lowest processing time and highest accuracy	(92)
2	Recall = 0.93, Precision = 0.871 with lower processing time	The system is cost-effective by reducing processing time and sustainable by increasing the accuracy values considerably. The proposed framework can also be used to prioritize patients who require an ambulance.	(93)
3	Accuracy = 93% and recall score = 88% using chest x-ray images	The proposed method can successfully help radiologist's prompt detection of coronavirus cases	(94)
4	Accuracy (97.94 %) and AUC (98.39 %)	A channel-shuffled dual-branched CNN architecture can effectively learn salient features and increases the accuracy and precision values of the modeling	(95)
5	Sensitivity = 100%, specificity = 99.02% and accuracy = 99.51% and for radiology data, sensitivity 89.21%, specificity = 83.33% and accuracy = 86.27%	This model is low cost and is used as a complementary method during C.T. imaging	(96)
6	Accuracy = 85.03%, sensitivity = 87.55%, specificity = 81.95%, precision = 85.01% and F1-core = 86.20%	Higher classification rate by analyzing thousands of images	(97)
7	Accuracy = 94.5%, confidence interval = 95%, sensitivity = 98.4% and specificity = 98.0%	Develops a DL-based CAD scheme of chest X-ray images and improves detecting COVID-19 infected	(40)
8	Accuracy = 98.97%, sensitivity = 89.39%, specificity = 99.75%, and an F-score = 96.72%	Reduces the misdiagnosis rates, and improves evaluation rates and detects positive COVID-19 infections	(98)
9	Accuracy = 97.9%, AUC = 99.0%, and Cohen kappa score = 95.7%.	Reliable screening of COVID-19 from chest CT	(99)
10	96% of accuracy	The proposed model performance is clinically validated with expert radiologists	(39)
11	Accuracy of 99.62 and 96.70%. Average recall value of 99.63 and 96.69%, respectively, for binary and multiclass	Automated medical diagnostics for enhancing decision making rates	(100)
12	Correlation coefficient = 0.9899	providing significant variance for each criterion	(101)
13	Accuracy = 99.7%, precision = 99.7%, and sensitivity = 99.7%	Improving the speed and accuracy of COVID-19 detection	(102)
14	86% accuracy for the task of classifying	The proposed model could successfully improve the classification accuracy	(103)
15	Accuracy of 88, 91, 87 and 89% for kNN, SVM, D.T. and L.R., respectively	The proposed method can be applied anywhere, without prior training or calibration	(104)
16	F1-score of 97.9, 98.8, and 92.5%, A.U.C. of 97.4, 98.8, and 84.4% and accuracy of 97, 98.2, and 88.2%, respectively, for Cough sound, Breathing sound and voices, respectively.	To improve the COVID-19 detection through a cost-effective approach	(105)
17	$R^2 = 0.96$, RMSE = 254, MAE = 186	The proposed method could successfully estimate the number of daily cases	(106)
18	Sensitivity and specificity of 0.8645, and 0.9889.	This model provides automated and accurate segmentation of C.T. images	(107)
19	MAPEs = 0.52, 0.38, 0.05, and 0.86%, respectively for the next 6 days in Wuhan, Beijing, Shanghai, and countrywide	To minimize the errors of the prediction and to enhance the detection efficiency	(108)
20	Accuracy = 98.84%, Precision = 93%, Sensitivity = 100%, and Specificity = 97.0%	The proposed model improved classification rate in comparison with ReseNet18, ReseNet50, Squeeze net, DenseNet-121, and Visual Geometry Group	(109)
21	Accuracy for both SVM and Decision Tree could provide the maximum value by average value of 93%	Higher accuracy for perceiving the perception of people infected by COVID-19	(110)
22	R.M.S.E. and CC values for five countries including, China, Italy, U.S.A., Iran and Germany	The proposed models enhanced the forecasting rate of COVID-19 cases	(91)
23	MAPE = 13.15% and CC = 0.99	The proposed models increased the forecasting rate of COVID-19 cases	(90)

DISCUSSION

According to the reviewed studies, the COVID-19 dataset can be imported from three primary sources: radiography, health centers' statistics, and Sensors for prediction, monitoring, identification, detection, diagnosis, and classification purposes. The output of the techniques needs to be evaluated to confirm the approach performance and accuracy values. The frequently used parameters for performance analysis include Accuracy, Precision, Recall, R.M.S.E., Correlation coefficient, and mean absolute

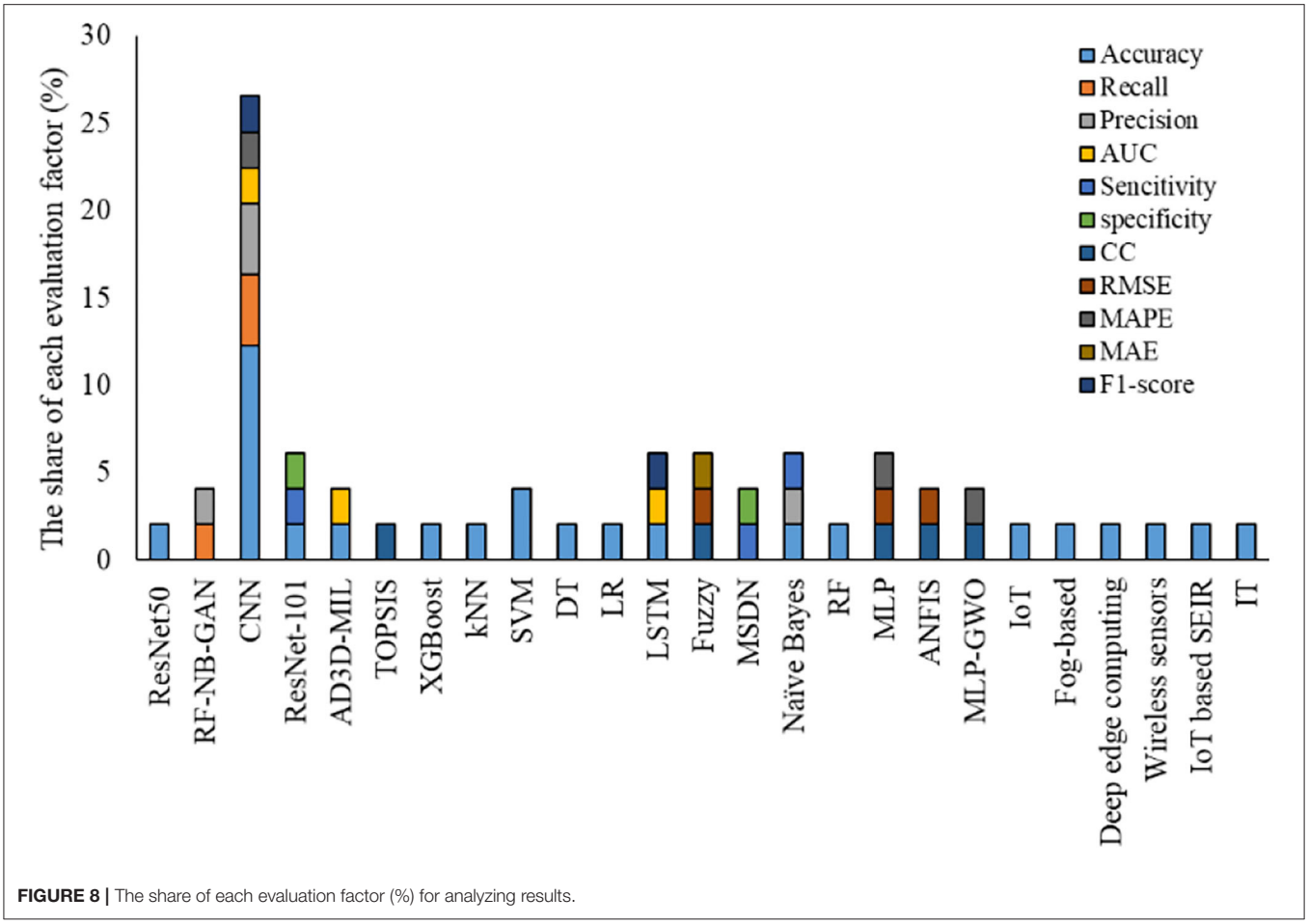


FIGURE 8 | The share of each evaluation factor (%) for analyzing results.

percentage error. This can be considered a brief explanation as the main contribution of the present study. This study successfully presents the advantages and disadvantages of each technique for a specific task in handling the COVID-19 dataset and proposes future perspectives. Also, this study can detect the main challenges and limitations.

It is also possible to find out which methodology is still available for which application can be considered a research opportunity for the future. Also, by carefully examining the different reasons for the tendency of each method to the fields shown in independent research, which can be considered necessary research and planning opportunities for policymakers in this field.

The presence of the ML platform led to reducing the adverse effects of the disease and accelerating the healing process, advances in treatment, medication, screening, prognosis, contact tracking, and the drug/vaccine development process, and reduced human intervention in medical performance as a tool for the management of virtual queues to prevent crowds in physical waiting rooms or long queues. It is used to predict waiting times and implement calls privately with the cell phone platform.

Based on the studies conducted in this study, we achieved the following results:

- IoT has been used more than other applications to monitor and detect COVID-19 cases. In contrast, it has been less popular in the identification.
- ML method is widely used in data analysis by producing an analytical model intelligently for estimating, categorizing, optimizing, predicting, identifying problems, and decision making.
- New computing technologies have made the problems assessed by ML-based techniques, began to evolve from pattern recognition to a comprehensive theory of the ability of computers to perform specific tasks without the need for special planning.
- Identifying the prevalence, effective parameters in eradicating the virus, identifying patients in the early stages, patients' pattern behaviors, and predicting outbreak and mortality rates can be considered practical and compelling areas of ML-based techniques.
- Detection, diagnosis, and prediction can be considered the main categories of the application of ML-based methods in COVID-19. In general, one of the main sections of analyzing IoT-based and ML-Based techniques applied for a specific field is their evaluation in terms of accuracy, error, or performance of the model.

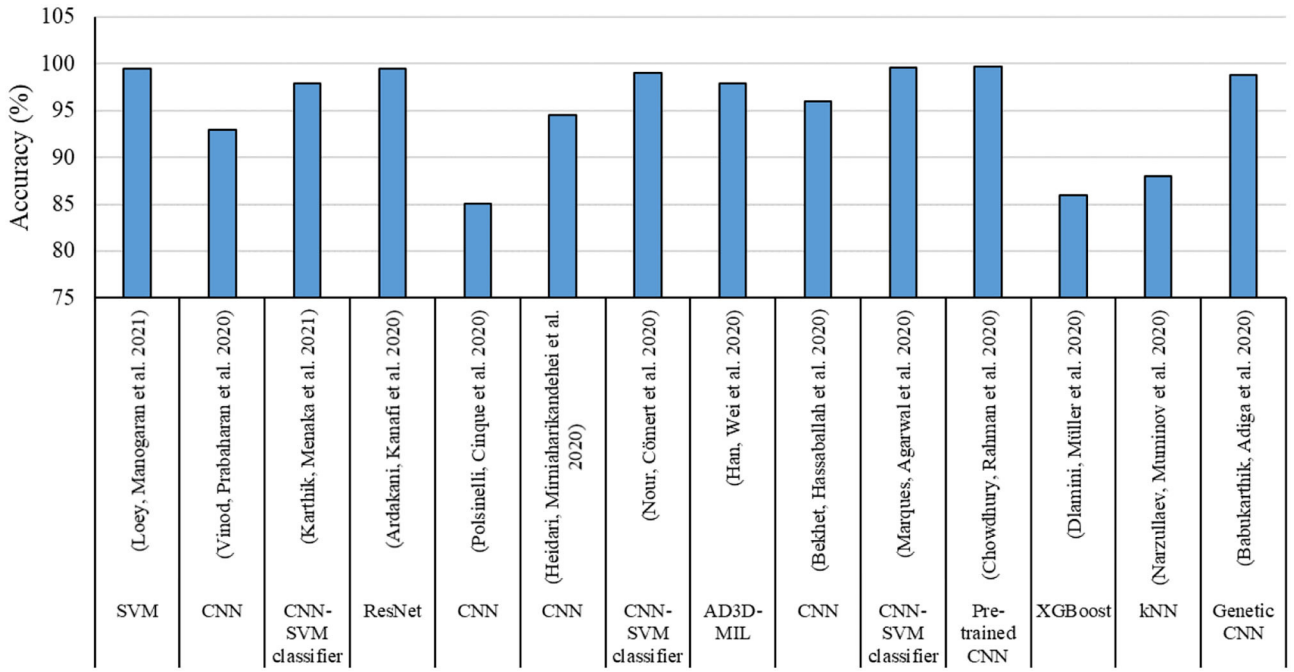


FIGURE 9 | Accuracy values.

- Accuracy, followed by the recall and precision parameters, has the highest portion of the evaluation criteria employed for analyzing the COVID-19 dataset using IoT and ML-based techniques. ResNet, as an architecture of deep learning methods followed by CNN, XGBoost, SVM, D.T., and L.R., has been used more often to tackle work with COVID-19 related data.
- Resnet follows CNN is The most common use of ML to contribute various methods for different tasks to combat Pandemic COVID-19. This trend can be due to the model's nature for handling different applications like monitoring, detection, identification, classification, and diagnosis. At the same time, other methods can do a limited number of applications.
- Models developed using ML and IoT-ML require an evaluation step for recognizing their performance and accuracy values. According to the studies reviewed, the most effective and frequently used evaluation criteria include Accuracy, Recall, Precision, Root mean square error (R.M.S.E.), Correlation coefficient, and Mean absolute percentage error (M.A.P.E.). These criteria compare the models' output and actual values and provide a comparison score (90, 91). In the present study, we employed the criteria values reported by each study for evaluating and comparing the models.
- Most of these studies lack numerical analysis for the method's performance. One of the main reasons can be the nature of the IoT technique, which goes through a practical process and shows its performance in practical applications and does not need to provide numerical statistics. In all these applications, IoT could successfully cope with the task. IoT provided a fast and efficient approach to tracking the disease spread. On the other hand, it can be employed as a real-time framework to minimize the impact of communicable diseases through the early detection of cases.
- The most share of studies developed by ML-based techniques for handling COVID-19 based dataset provided performance criteria. The most share of the performance criteria is related to the accuracy factor. The accuracy factor is general and normalized. Therefore, it can be employed for comparing the ML-based methods with different datasets.
- CNN, SVM classifier, Genetic CNN, and pre-trained CNN followed by ResNet provided the highest accuracy values. On the other hand, the lowest accuracy was related to single CNN, followed by XGboost and K.N.N. techniques.

CONCLUSION

The present study categorizes the applications of IoT, IoT-ML, and ML-based techniques to tackle COVID-19-related problems. The main applications are monitoring, detection, identification, classification, and diagnosis. Studying, comparing, and investigating these applications requires a proper judgment about the performance and effectiveness of outputs. According to a deep consideration of the evaluation criteria, it has been investigated that the accuracy, followed by the recall and precision parameters, have owned the highest portion of the evaluation criteria employed for analyzing COVID-19 based dataset using IoT and ML-based techniques.

Most of the studies lack numerical analysis for the method performance. One of the main reasons can be the nature of the IoT technique which goes through a practical process and shows its performance in practical applications. In all the applications, IoT could successfully cope with the tasks. Such that, IoT provided a fast and efficient approach to tracking the disease spread. Most of the studies developed by ML-based techniques for handling COVID-19-based datasets provided performance criteria. According to the results section, the following points can be extracted:

- IoT provided a fast and efficient approach to tracking the disease spread.
- IoT can be employed as a real-time framework to minimize the impact of communicable diseases through early detection of cases.
- The most popular performance criteria are related to the accuracy factor.
- ML-based methods are able to be used with different types of datasets.
- CNN with SVM classifier, Genetic CNN, and pre-trained CNN followed by ResNet, provided the highest accuracy values.
- A.I. is a result-oriented technology employed for proper screening, analysis, forecasting, and tracking of current and potential future patients.

Policy-making in COVID-19 disease to examine the weaknesses and strengths and vulnerabilities of society in terms of the penetration of pathogenic viruses can be considered additional measures and future studies. On the other hand, the study of collective behaviors can also be considered as a perspective to complete studies to prevent similar social harms, reduce costs incurred, and not surprise human life. The future perspective is to employ an advanced analytic ML-based platform that supports huge-data analytics. This trend moves toward smart health interconnected with innovative technologies in the sensor industry. The future is waiting for tremendous promotion in smart health.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SB and SA designed the study. SA, AY, BP, AK, and AM wrote the paper. SB, AB, HA-R, MM, and AK edited the manuscript. SA, AY, and BP carried out all the analyses. SB, SA, AY, BP, and MM generated all figures and all tables. HA-R and AB were not involved in any analyses. All authors have read and approved the final version of the paper.

REFERENCES

1. Yuki K, Fujiogi M, Koutsogiannaki S. COVID-19 pathophysiology: a review. *Clin Immunol.* (2020) 215:108427. doi: 10.1016/j.clim.2020.108427
2. Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov.* (2020) 19:305–6. doi: 10.1038/d41573-020-00073-5
3. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nat Hum Behav.* (2020) 4:1303–12. doi: 10.1038/s41562-020-01009-0
4. Marshall JC, Murthy S, Diaz J, Adhikari N, Angus DC, Arabi YM, et al. A minimal common outcome measure set for COVID-19 clinical research. *Lancet Infect Dis.* (2020) 20:e192–7. doi: 10.1016/S1473-3099(20)30483-7
5. Cao X. COVID-19: immunopathology and its implications for therapy. *Nat Rev Immunol.* (2020) 20:269–70 doi: 10.1038/s41577-020-0308-3
6. Ciotti M, Ciccozzi M, Terrinoni A, Jiang WC, Wang CB, Bernardini S. The COVID-19 pandemic. *Crit Rev Clin Lab Sci.* (2020) 57:365–88. doi: 10.1080/10408363.2020.1783198
7. Mahmoudi MR, Heydari MH, Qasem SN, Mosavi A, Band SS. Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries. *Alexandria Eng J.* (2021) 60:457–64. doi: 10.1016/j.aej.2020.09.013
8. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic. *Diabetes Metab Syndr.* (2020) 14:569–73. doi: 10.1016/j.dsx.2020.05.008
9. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr.* (2020) 14:337–9. doi: 10.1016/j.dsx.2020.04.012
10. Mendhe D, Wankhede P, Wanjari M, Alwadkar S. Mucormycotic osteomyelitis of maxilla post-COVID patient: a case report. *Pan Afr Med J.* (2021) 39:275. doi: 10.11604/pamj.2021.39.275.30480
11. Ahmadinia M, Ahangarikiasari H. Data aggregation in wireless sensor networks based on environmental similarity: a learning automata approach. *J Netw.* (2014) 9:2567. doi: 10.4304/jnw.9.10.2567-2573
12. Ahmadinia M, Meybodi MR, Esnaashari M. Energy-efficient and multi-stage clustering algorithm in wireless sensor networks using cellular learning automata. *IETE J Res.* (2013) 59:774–82. doi: 10.4103/0377-2063.126958
13. Alinejad-Rokny H, Pourshaban H, Orimi AG. Network motifs detection strategies and using for bioinformatic networks. *J Bionanosci.* (2014) 8:353–9. doi: 10.1166/jbns.2014.1245
14. Bahrani P, Minaei-Bidgoli B, Parvin H, Mirzarezaee M, Keshavarz A. User and item profile expansion for dealing with cold start problem. *J Intell Fuzzy Syst.* (2020) 38:4471–83. doi: 10.3233/JIFS-191225
15. Mahmoudi MR, Akbarzadeh H, Parvin H, Nejatian S, Rezaie V. Consensus function based on cluster-wise two level clustering. *Artif Intell Rev.* (2021) 54:639–65. doi: 10.1007/s10462-020-09862-1
16. Niu H, Khozouie N, Parvin H, Beheshti A. An ensemble of locally reliable cluster solutions. *Appl Sci.* (2020) 10:1891. doi: 10.3390/app10051891
17. Niu H, Xu W, Akbarzadeh H, Parvin H, Beheshti A. Deep feature learnt by conventional deep neural network. *Comput Elect Eng.* (2020) 84:106656. doi: 10.1016/j.compeleceng.2020.106656
18. Parvin, H. Behrouz Minaei, Hosein Alizadeh, and Akram Beigi. A novel classifier ensemble method based on class weightening in huge dataset. In: *International Symposium on Neural Networks.* (2011). p. 144–150
19. Parvin H, Parvin S. A classifier ensemble of binary classifier ensembles. *Int J Learn Manage Syst.* (2013) 1:37–47. doi: 10.12785/ijlms/010204
20. Parvin H. MirnabiBaboli, M. Proposing a classifier ensemble framework based on classifier selection and decision tree. *Eng Appl Artif Intell.* (2015) 37:34–42. doi: 10.1016/j.engappai.2014.08.005
21. Alinejad-Rokny H, Pedram MM. Discovered motifs with using parallel Mprefixspan method. *Sci Res Essays.* (2011) 6:4220–6. doi: 10.5897/SRE11.212

22. Parvin H, Seyedaghaee N. A heuristic scalable classifier ensemble of binary classifier ensembles. *J Bioinform Intell Control.* (2012) 1:163–70. doi: 10.1166/jbic.2013.1016
23. Alinejad-Rokny, H. Proposing on optimized homolographic motif mining strategy based on parallel computing for complex biological networks. *J Med Imag Health Inform*. (2016) 6:416–24. doi: 10.1166/jmihi.2016.1707
24. Alinejad-Rokny H, Anwar F, Waters SA, Davenport MP. Source of CpG depletion in the HIV-1 genome. *Mol Biol Evol.* (2016) 33:3205–12. doi: 10.1093/molbev/msw205
25. Alinejad-Rokny H, Sadroddiny E. Machine learning and data mining techniques for medical complex data analysis. *Neurocomputing.* (2018) 276. doi: 10.1016/j.neucom.2017.09.027
26. Bayati M, Rabiee HR, Mehrbod M, Vafaee F, Ebrahimi D, Forrest AR. CANCERSIGN: a user-friendly and robust tool for identification and classification of mutational signatures and patterns in cancer genomes. *Sci Rep.* (2020) 10:1–11. doi: 10.1038/s41598-020-58107-2
27. Dashti H, Dehzangi A, Bayati M, Breen J, Lovell N, Ebrahimi D. Integrative analysis of mutated genes and mutational processes reveals seven colorectal cancer subtypes. *BMC Bioinform.* (2022) 23:1–24. doi: 10.1186/s12859-022-04652-8
28. Ghareyazi A, Mohseni A, Dashti H, Beheshti A, Dehzangi A, Rabiee HR. Whole-genome analysis of de novo somatic point mutations reveals novel mutational biomarkers in pancreatic cancer. *Cancers.* (2021) 13:4376. doi: 10.3390/cancers13174376
29. Heidari R, Akbariqomi M, Asgari Y, Ebrahimi D. A systematic review of long non-coding RNAs with a potential role in Breast Cancer. *Mut Res Rev Mut Res.* (2021) 787:108375. doi: 10.1016/j.mrrev.2021.108375
30. Javanmard R, JeddiSaravi K. Proposed a new method for rules extraction using artificial neural network and artificial immune system in cancer diagnosis. *J Bionanosci.* (2013) 7:665–72. doi: 10.1166/jbns.2013.1160
31. Kalantari AK. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions. *Neurocomputing.* (2018) 276:2–22. doi: 10.1016/j.neucom.2017.01.126
32. Shamshirband S, Fathi M, Dehzangi A, Chronopoulos AT. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J Biomed Inform.* (2021) 113:103627. doi: 10.1016/j.jbi.2020.103627
33. Shamshirband S, Rabczuk T, Chau KW. A survey of deep learning techniques: application in wind and solar energy resources. *IEEE Access.* (2019)7:164650–66. doi: 10.1109/ACCESS.2019.2951750
34. Khakmardan S, Rezvani M, Pouyan AA, Fateh M. MHiC, an integrated user-friendly tool for the identification and visualization of significant interactions in Hi-C data. *BMC Genom.* (2020) 21:1–10. doi: 10.1186/s12864-020-6636-7
35. Rajaei P, Jahanian KH, Beheshti A, Band SS, Dehzangi A. VIRMOTIF: a user-friendly tool for viral sequence analysis. *Genes.* (2021) 12:186. doi: 10.3390/genes12020186
36. Gooneratne SL, Alinejad-Rokny H, Ebrahimi D, Bohn PS, Wiseman RW, O'Connor DH, et al. Linking pig-tailed macaque major histocompatibility complex class I haplotypes and cytotoxic T lymphocyte escape mutations in simian immunodeficiency virus infection. *J Virol.* (2014) 88:14310–25. doi: 10.1128/JVI.02428-14
37. Sharifrazi D, Alizadehsani R, Joloudari JH, Shamshirband S, Hussain S, Sani ZA. CNN-KCL: Automatic myocarditis diagnosis using convolutional neural network combined with k-means clustering. *Preprint.* (2020). doi: 10.20944/preprints202007.0650.v1
38. Aman AH, Hassan WH, Sameen S, Attarbashi ZS, Alizadeh M, Latiff LA, et al. IoMT amid COVID-19 pandemic: Application, architecture, technology, and security. *J Netw Comput Appl.* (2020) 174:102886. doi: 10.1016/j.jnca.2020.102886
39. Bekhet S, Hassaballah M, Kenk MA, Hameed MA. *An Artificial Intelligence Based Technique for COVID-19 Diagnosis from Chest X-Ray.* (2020). p. 191–5.
40. Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *Int J Med Inform.* (2020) 144:104284. doi: 10.1016/j.ijmedinf.2020.104284
41. Sufian A, Jat DS, Banerjee A, Sufian A, Jat DS, Banerjee A (2020). Insights of artificial intelligence to stop spread of covid-19. In: *Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach*. Cham: Springer. p. 177–90. doi: 10.1007/978-3-030-55258-9_11
42. Barstugan M, Ozkaya U, Ozturk S. Coronavirus (covid-19) classification using ct images by machine learning methods. *arXiv preprint*. arXiv:2003.09424 (2020)
43. Özkaya U, Öztürk S, Budak S, Melgani F, Polat K. Classification of COVID-19 in chest CT images using convolutional support vector machines. *arXiv preprint*. arXiv:2011.05746 (2020)
44. Öztürk S, Özkaya U, Barstugan M. Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features. *Int J Imaging Syst Technol.* (2021) 31:5–15. doi: 10.1002/ima.22469
45. Guo X, Zhang YD, Lu S, Lu Z. A survey on machine learning in COVID-19 diagnosis. *CMES-Computer Modeling in Engineering and Sciences.* (2021). doi: 10.32604/cmes.2022.017679
46. Abumalloh RA, Nilashi M, Yousoof Ismail M, Alhargan A, Alghamdi A, Alzahrani AO, et al. Medical image processing and COVID-19: a literature review and bibliometric analysis. *J Infect Public Health.* (2022) 15:75–93. doi: 10.1016/j.jiph.2021.11.013
47. Khan M, Mehran MT, Haq ZU, Ullah Z, Naqvi SR, Ihsan M, et al. Applications of artificial intelligence in COVID-19 pandemic: a comprehensive review. *Exp Syst Appl.* (2021) 185:115695. doi: 10.1016/j.eswa.2021.115695
48. El-Rashidy N, Abdelrazik S, Abuhmed T, Amer E, Ali F, Hu J-W, et al. Comprehensive survey of using machine learning in the COVID-19 pandemic. Diagnostics (Basel). (2021) 11:1155. doi: 10.3390/diagnostics11071155
49. Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: a review. *Inform Med Unlocked.* (2021) 24:100564. doi: 10.1016/j.imu.2021.100564
50. Koutsos TM, Menexes GC, Dordas CA. An efficient framework for conducting systematic literature reviews in agricultural sciences. *Sci Total Environ.* (2019) 682:106–17. doi: 10.1016/j.scitotenv.2019.04.354
51. Fleming PS, Koletsi D, Pandis N. Blinded by PRISMA: are systematic reviewers focusing on PRISMA and ignoring other guidelines? *PLoS ONE.* (2014) 9:e96407. doi: 10.1371/journal.pone.0096407
52. Page MJJE, McKenzie PM, Bossuyt I, Boutron TC, Hoffmann CD, Mulrow L, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* (2021) 372.
53. Bandyopadhyay D, Sen J. Internet of things: applications and challenges in technology and standardization. *Wire Pers Commun.* (2011) 58:49–69. doi: 10.1007/s11277-011-0288-5
54. Zeinab KA, Elmustafa SA. Internet of things applications, challenges and related future technologies. *World Sci News.* (2017) 67:126–48.
55. Keramidas G, Voros N, Hübner M. *Components and Services for IoT Platforms*. Cham: Springer. (2016). doi: 10.1007/978-3-319-42304-3
56. Adryan B, Obermaier D, Fremantle P. *The Technical Foundations of IoT*. Washington, DC: Artech House. (2017).
57. Nagaraj. *Introduction to Sensors in IoT and Cloud Computing Applications*. Potomac: Bentham Science Publishers (2021).
58. Rahman MA, Asyhari AT. The emergence of Internet of Things (IoT): connecting anything, anywhere. *Computers.* (2019) 8:40. doi: 10.3390/computers8020040
59. Boumehrez F, Sahour AH, Doghmane N. Telehealth care enhancement using the internet of things technology. *Bull Electr Eng Inform.* (2021) 10:2652–60. doi: 10.11591/eei.v10i5.2968
60. Bolhasani H, Mohseni M, Rahmani AM. Deep learning applications for IoT in health care: a systematic review. *Inform Med Unlocked.* (2021) 23:100550. doi: 10.1016/j.imu.2021.100550
61. Marques G, Bhoi AK, Victor Hugo C, Hareesha K. *IoT in Healthcare and Ambient Assisted Living.* Cham: Springer. (2021). doi: 10.1007/978-981-15-9897-5
62. Chamola V, Hassija V, Gupta V, Guizani M, A. comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. *IEEE Access.* (2020) 8:90225–65. doi: 10.1109/ACCESS.2020.2992341
63. Taylor W, Abbasi QH, Dashtipour K, Ansari S, Shah SA, Khalid A, et al. Review of the state of the art in non-contact sensing for COVID-19. *Sensors.* (2020) 20:5665. doi: 10.3390/s20195665

64. Khan MB, Zhang Z, Li L, Zhao W, Hababi MAMA, Yang X, et al. A systematic review of non-contact sensing for developing a platform to contain COVID-19. Micromachines. (2020) 11:912. doi: 10.3390/mi11100912
65. Mohammed B, Isa SM. The role of internet of things (IoT) in the containment and spread of the novel COVID-19 pandemic. In: *Computational Intelligence Methods in COVID-19: Surveillance, Prevention, Prediction and Diagnosis*. Cham: Springer (2021). p. 109–19. doi: 10.1007/978-981-15-8534-0_6
66. Roy A, Kumbhar FH, Dhillon HS, Saxena N, Shin SY, Singh S. Efficient monitoring and contact tracing for COVID-19: a smart IoT-based framework. *IEEE Internet Things Mag.* (2020) 3:17–23. doi: 10.1109/IOTM.0001.2000145
67. Rahman A, Hossain MS, Alrajeh NA, Alsolami F. Adversarial examples—Security threats to COVID-19 deep learning systems in medical IoT devices. *IEEE Int Things J.* (2020) 8:9603–10. doi: 10.1109/JIOT.2020.3013710
68. Ahmed I, Ahmad A, Jeon G. An IoT-based deep learning framework for early assessment of COVID-19. *IEEE Internet Things J.* (2020) 8:15855–62 doi: 10.1109/JIOT.2020.3034074
69. Otoom M, Otoum N, Alzubaidi MA, Etoom Y, Banihani R. An IoT-based framework for early identification and monitoring of COVID-19 cases. *Biomed Signal Process Control.* (2020) 62:102149. doi: 10.1016/j.bspc.2020.102149
70. Singh RP, Javaid M, Haleem A, Suman R. Internet of things (IoT) applications to fight against COVID-19 pandemic. *Diabetes Metab Syndr Clin Res Rev.* (2020) 14:521–4. doi: 10.1016/j.dsx.2020.04.041
71. Vedaei SS, Fotovvat A, Mohebbian MR, Rahman GM, Wahid KA, Babyn P, et al. an IoT-based system for automated health monitoring and surveillance in post-pandemic life. *IEEE Access.* (2020) 8:188538. doi: 10.1109/ACCESS.2020.3030194
72. Ashraf MU, Hannan A, Cheema SM, Ali Z, Alofi A. *Detection and Tracking Contagion using IoT-Edge Technologies: Confronting COVID-19 Pandemic.* (2020). p. 1–6.
73. Baskaran P, Rajaram B, Kumaratharan N. *IoT Based COVID Preventive System for Work Environment.* (2020). p. 65–71. doi: 10.1109/I-SMAC49090.2020.9243471
74. Karmore S, Bodhe R, Al-Turjman F, Kumar RL, Pillai S. IoT based humanoid software for identification and diagnosis of Covid-19 suspects. *IEEE Sens J.* (2020) doi: 10.1109/JSEN.2020.3030905
75. Akhund TMNU, Jyoty WB, Siddik MAB, Newaz NT, Al Wahid SA, Sarker MM. *IoT Based Low-Cost Robotic Agent Design for Disabled and Covid-19 Virus Affected People.* (2020). p. 23–6.
76. Ndiaye M, Oyewobi SS, Abu-Mahfouz AM, Hancke GP, Kurien AM, Djouani K. IoT in the wake of COVID-19: a survey on contributions, challenges and evolution. *IEEE Access.* (2020) 8:186821–39. doi: 10.1109/ACCESS.2020.3030090
77. Wang B, Sun Y, Duong TQ, Nguyen LD, Hanzo L. Risk-aware identification of highly suspected COVID-19 cases in social IoT: a joint graph theory and reinforcement learning approach. *IEEE Access.* (2020) 8:115655–61. doi: 10.1109/ACCESS.2020.3003750
78. Kumar K, Kumar N, Shah R. Role of IoT to avoid spreading of COVID-19. *Int J Intell Netw.* (2020) 1:32–5. doi: 10.1016/j.ijin.2020.05.002
79. Kolhar M, Al-Turjman F, Alameen A, Abualhaj MM, A. three layered decentralized IoT biometric architecture for city lockdown during COVID-19 outbreak. *IEEE Access.* (2020) 8:163608–17. doi: 10.1109/ACCESS.2020.3021983
80. De Vito S, Esposito E, D'Elia G, Del Giudice A, Fattoruso G, Ferlito S, et al. *High Resolution Air Quality Monitoring with IoT Intelligent Multisensor devices during COVID-19 Pandemic Phase 2 in Italy.* (2020). p. 1–6.
81. Manalu EP, Muditomo A, Adriana D, Trisnowati Y. Role *of Information Technology for Successful Responses to Covid-19 Pandemic.* (2020). p. 415–20.
82. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: a review. *Chaos Solitons Fractals.* (2020) 139:110059. doi: 10.1016/j.chaos.2020.110059
83. Kyritsis I, Deriaz M. *A Queue Management Approach for Social Distancing and Contact Tracing.* (2020). p. 66–8. doi: 10.1109/AI4I49448.2020.00022
84. Jordan I, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science.* (2015) 349:255–60. doi: 10.1126/science.aaa8415
85. Sra S, Nowozin S, Wright SJ. *Optimization for Machine Learning.* Cambridge: Mit Press. (2012).
86. Kashyap P. *Machine Learning for Decision Makers: Cognitive Computing Fundamentals for Better Decision Making.* Berlin: Springer. (2017). doi: 10.1007/978-1-4842-2988-0
87. Abdulkareem H, Mohammed MA, Gunasekaran SS, Al-Mhiqani MN, Mutlag AA, Mostafa SA, et al. A review of fog computing and machine learning: concepts, applications, challenges, and open issues. *IEEE Access.* (2019) 7:153123–40. doi: 10.1109/ACCESS.2019.2947542
88. Shinde PP, Shah S. *A Review of Machine Learning and Deep Learning Applications.* (2018). p. 1–6.
89. Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ digital medicine.* (2018) 1:1–3. doi: 10.1038/s41746-018-0061-1
90. Ardabili S, Mosavi A, Band SS, Varkonyi-Koczy AR. *Coronavirus Disease (COVID-19) Global Prediction Using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer.* (2020). p. 000251–4 doi: 10.34055/osf.io/vfcex
91. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. Covid-19 outbreak prediction with machine learning. *Algorithms.* (2020) 13:249. doi: 10.3390/a13100249
92. Loey GM. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement.* (2021) 167:108288. doi: 10.1016/j.measurement.2020.108288
93. Singh P, Kaur R. An integrated fog and Artificial Intelligence smart health framework to predict and prevent COVID-19. *Global Trans.* (2020) 2:283–92. doi: 10.1016/j.glt.2020.11.002
94. Vinod N, Prabaharan SJC. Solitons, and Fractals. Data science and the role of Artificial Intelligence in achieving the fast diagnosis of Covid-19. *Chaos Solitons Fractals.* (2020) 140:110182. doi: 10.1016/j.chaos.2020.110182
95. Karthik R, Menaka R, Hariharan M. Learning distinctive filters for COVID-19 detection from chest X-ray using shuffled residual CNN. *Appl Soft Comput.* (2021) 99:106744. doi: 10.1016/j.asoc.2020.106744
96. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput Biol Med.* (2020) 121, 103795. doi: 10.1016/j.compbiomed.2020.103795
97. Polsinelli M, Cinque L, Placidi G, A. light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognit Lett.* (2020) 140:95–100. doi: 10.1016/j.patrec.2020.10.001
98. Nour M, Cömert Z, Polat K, A. novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Appl Soft Comput.* (2020) 97:106580. doi: 10.1016/j.asoc.2020.106580
99. Han Z, Wei B, Hong Y, Li T, Cong J, Zhu X, et al. Accurate screening of COVID-19 using attention-based deep 3D multiple instance learning. *IEEE Trans Med Imaging.* (2020) 39:2584–94. doi: 10.1109/TMI.2020.2996256
100. Marques G, Agarwal D. de la Torre Díez I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl Soft Comput.* (2020) 96:106691. doi: 10.1016/j.asoc.2020.106691
101. Mohammed MA, Abdulkareem KH, Al-Waisy AS, Mostafa SA, Al-Fahdawi S, Dinar AM, et al. Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and TOPSIS methods. *IEEE Access.* (2020) 8:99115–31. doi: 10.1109/ACCESS.2020.2995597
102. Chowdhury ME, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub Z B, et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access.* (2020) 8:132665–76. doi: 10.1109/ACCESS.2020.3010287
103. Dlamini GS, Müller SJ, Meraba RL, Young RA, Mashiyane J, Chiwewe T, et al. Classification of COVID-19 and other pathogenic sequences: a dinucleotide frequency and machine learning approach. *IEEE Access.* (2020) 8:195263–73. doi: 10.1109/ACCESS.2020.3031387
104. Narzullaev A, Muminov Z, Narzullaev M. *Contact Tracing of Infectious Diseases Using Wi-Fi Signals and Machine Learning Classification.* (2020). p. 1–5.
105. Hassan A, Shahin I, Alsabek MB. Covid-19 detection system using recurrent neural networks. (2020). p. 1–5.
106. Cihan P. Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak. (2020). p. 1–4.
107. Zheng B, Liu Y, Zhu Y, Yu F, Jiang T, Yang D, et al. MSD-Net: multi-scale discriminative network for COVID-19 lung infection segmentation on CT. *IEEE Access.* (2020) 8:185786–95. doi: 10.1109/ACCESS.2020.3027738

108. Zheng N, Du S, Wang J, Zhang H, Cui W, Kang Z, et al. Predicting COVID-19 in China using hybrid AI model. *IEEE Transac Cybern*. (2020) 50:2891–904. doi: 10.1109/TCYB.2020.2990162
109. Babukarthik R, Adiga VAK, Sambasivam G, Chandramohan D, Amudhavel JJIA. Prediction of COVID-19 Using Genetic Deep Learning Convolutional Neural Network (GDCNN). *IEEE Access*. (2020) 8:177647–66. doi: 10.1109/ACCESS.2020.3025164
110. Sethi M, Pandey S, Trar P, Soni P. Sentiment identification in COVID-19 specific tweets. (2020). p. 509–16. doi: 10.1109/ICESC48915.2020.9155674

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Band, Ardabili, Yarahmadi, Pahlevanzadeh, Kiani, Beheshti, Alinejad-Rokny, Dehzangi, Chang, Mosavi and Moslehpour. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fpubh.2022.941958

Mechanism of Negative Emotions of the Elderly in Normalization Period of COVID-19: A Mediated Mediation Model

*Kai Xu**

School of Educational Science, Luoyang Normal University, Luoyang, China

To explore the mechanism of negative emotions of the elderly in normalization period of COVID-19. The self-rating Depression Anxiety Stress Scale, epidemic attention scale, subjective economic status scale and physical health perception scale were used to investigate 318 elderly people in 2021. There were significant differences in negative emotions among the elderly in Henan in China with different gender, education background, medical insurance and whether they contacted suspected cases (all $P < 0.05$), but there was no significant difference on religious belief ($P > 0.05$); Attention to epidemic information was positively correlated with negative emotion ($r = 0.492$, $P < 0.01$), and negatively correlated with subjective economic status ($r = -0.138$, $P < 0.05$); Negative emotions were negatively correlated with subjective economic status ($r = -0.455$, $P < 0.01$) and health perception ($r = -0.277$, $P < 0.01$); health perception was no significant correlation with epidemic attention($r = -0.047$, $P > 0.05$) and subjective economic status ($r = -0.033$, $P > 0.05$). Bootstrap test found that epidemic attention can significantly predict negative emotion of the elderly ($\beta = 0.492$, $P < 0.001$), subjective economic status played a partial mediating role between epidemic attention and negative emotions ($\beta = 0.438$, -0.395, $P < 0.001$), and health perception moderated the first half of the mediating path ($\beta = 0.403$, $P < 0.001$, 95% CI = [0.286∼0.521]). Epidemic attention has a significant positive impact on the negative emotions of the elderly in Henan during normalization period of COVID-19, and it has effect indirectly through subjective economic status; health perception plays a moderator role in the impact of epidemic attention on subjective economic status.

Keywords: the elderly, epidemic attention, negative emotion, subjective economic status, health perception

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Parameshachari B. D., Visvesvaraya Technological University, India
Zhiwei Guo, Chongqing Technology and Business University, China
Xin Qi, Waseda University, Japan

***Correspondence:**
Kai Xu
xukai@lynu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *12 May 2022*
Accepted: *08 June 2022*
Published: *28 June 2022*

Citation:
Xu K (2022) Mechanism of Negative Emotions of the Elderly in Normalization Period of COVID-19: A Mediated Mediation Model. Front. Public Health 10:941958. doi: 10.3389/fpubh.2022.941958

INTRODUCTION

As a major public health emergency, COVID-19 is characterized by multiple transmission routes, long incubation period, strong infectivity and rapid spread. It is difficult to prevent and control the epidemic. It not only threatens people's life and physical health, but also brings various impacts on people's mental health. Due to the degenerative changes of various physical and mental functions, the elderly have become a vulnerable group in the society, which is prone to negative emotional experience, resulting in serious psychological problems (1). In this epidemic, due to the weak immune function, the elderly have become the susceptible group and high-risk vulnerable group

of this infectious disease, and most of the elderly in the critically ill group. The study also found that during the epidemic period, the incidence of depression and anxiety in the elderly increased significantly (2), and they were more prone to poor mental health and anxiety symptoms (3), which was the focus of social psychological counseling and crisis intervention during the epidemic period. At present, there are few studies on the mental health of the elderly in Chinese normalization period of COVID-19, which mainly focusing on the current situation and influencing factors of the frailty of the elderly in the community and the health management strategy of the elderly with chronic diseases (4, 5). Therefore, how to understand the law of behavior and emotion of the elderly in time and guide them to respond to public health emergencies rationally and scientifically has become an urgent topic to be studied.

In major emergencies such as the epidemic, people's mental health problems are mainly manifested in three negative emotions: stress, depression and anxiety (6). At present, there have been some researches on the mental health of the elderly during the epidemic period, but there are still deficiencies in the influence mechanism of the negative emotion of the elderly in China during the epidemic period, which needs to be further clarified. The important factor affecting people's mental health in the early stage of the epidemic is the epidemic information (7), people's cognitive uncertainty exists for a long time (8), and they tend to process negative information, which will lead to various serious negative emotional problems (9). Therefore, this study puts forward hypothesis 1:

> Epidemic concern has a significant positive impact on the negative emotions of the elderly.

Socio economic status is an important concept in the theory of social stratification. Socio economic status is the position of an individual in society and the comprehensive embodiment of his education level, occupation, income, residence and other indicators (10). The research on China's national mental health literacy found that in terms of social influencing factors, socio-economic status is the most effective factor among all variables, and its interpretation rate of Chinese mental health literacy is the highest, that is, the higher the socio-economic status, the higher the mental health literacy (11). Socio economic status is often divided into objective socio-economic status and subjective socio-economic status. The latter refers to the evaluation of their socio-economic status by the elderly compared with their peers. In the impact on the mental health of the elderly, the subjective economic status has an important impact. The higher the social class positioning, the better the mental health of the elderly (12). Therefore, this study puts forward hypothesis 2:

> Subjective economic status has a significant negative impact on the negative emotion of the elderly, and plays an intermediary role between epidemic concern and negative emotion.

Physical health perception is an individual's subjective perception and evaluation of self-health. Physical health perception level is a good indicator of health status. It can not only reflect individual health status, but also predict mortality (13). In the face of all kinds of information about the COVID-19, the elderly are more sensitive than other people. Their bodies will release a large number of signals, resulting in autonomic nervous symptoms such as elevated blood pressure, panic, suffocation, asthma, sweating, and various physical discomfort (14). The sense of physical health is closely related to negative emotions. Individuals with lower physical health perception have more negative emotions, and individuals with better physical health perception have more positive emotions (15). It can be seen that under the same conditions of epidemic concern and subjective economic status, the elderly with low physical perception will have more negative emotions, and the elderly with high physical health perception will have less negative emotions. Therefore, this study puts forward hypothesis 3:

> Physical health perception plays a regulatory role in the impact of epidemic concern on subjective economic status.

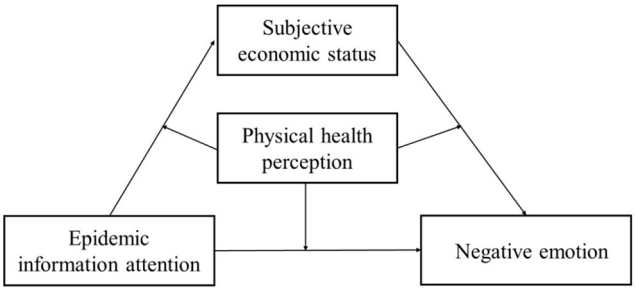


FIGURE 1 | Hypothetical model diagram of this study.

To sum up, this study proposes a moderated mediation model for the impact mechanism of negative emotions of the elderly during the COVID-19 period, as shown in **Figure 1**. Specifically, it includes the intermediary role of subjective economic status between epidemic concern and negative emotion, as well as the regulatory role of physical health perception in the direct path and intermediary path, so as to provide scientific support and thinking for effective psychological counseling and assistance to the elderly in the period of epidemic normalization.

RESEARCH OBJECT AND METHOD

Research Object

Using the method of convenient sampling, 318 elderly people were investigated in Henan Province from March to May 2021. Inclusion criteria: (1) age $\geq$ 55 years old; (2) Be conscious and able to answer questions independently; (3) Informed consent to this questionnaire and voluntary acceptance of this study. Exclusion criteria: (1) those who refuse to investigate; (2) Those with cognitive and language barriers cannot accurately understand the questionnaire items; (3) People with serious physical and mental diseases. The questionnaire adopts unified guiding language, and the elderly in their families

are investigated by Henan students from their departments and teaching departments. Since face-to-face investigation cannot be conducted during the epidemic, the online structured self-filling electronic questionnaire is adopted. After obtaining the informed consent of the elderly, the questionnaire is distributed by sharing the QR code and link of the questionnaire, which is filled in by the subjects themselves. The study passed the review of ethics of Luoyang Normal University.

Research Tool

(1) Demographic characteristics questionnaire: The contents of the questionnaire include gender, age, educational background, religious belief, whether there is medical insurance, whether there is contact with suspected cases, etc.
(2) Measurement of negative emotions: The self-rating Depression Anxiety Stress Scale (DASS-21) was developed by Lovibond et al. (16). Anxiety, depression and stress are three typical negative emotional experiences, and they are also the research hotspots of mental health in theory and clinic. Accurate and rapid diagnosis and differentiation of individual negative emotional status is the prerequisite to ensure individual mental health (17). Therefore, this study uses DASS-21 to measure mental health. Studies have confirmed that DASS-21 has good reliability and validity in the Chinese elderly (18–20). There are 21 items in the scale, which is divided into three subscales: depression, anxiety and stress. Each subscale contains 7 items. This scale adopts a 5-level score to allow the subjects to judge the consistency between their recent negative emotional experience or corresponding physiological response and the description of each item. The higher the score, the lower the level of mental health (21, 22). Diagnostic criteria for DASS-21 scores are detailed in **Table 1**.
(3) This paper adopts the epidemic concern questionnaire (23), which includes a question item ("how much do you pay attention to the epidemic situation of coronavirus pneumonia recently?"). The questionnaire adopts 5-level scoring. The higher the score, the higher the degree of concern about the epidemic.
(4) Subjective economic status scale: We used the MacArthur scale of subjective SES (24, 25) to present a 5-step ladder to the subjects, which means that the economic income is getting higher and higher from bottom to top. The subjects were asked to self-evaluate that their family economic income was at the level of the ladder, so as to represent their perceived subjective economic status.
(5) Physical health perception scale: Based on the overall physical health perception (HPQ) scale of HEO et al. (26), our survey of the elderly in China has good reliability and validity (27). The scale includes 4 items, e.g., "How would you rate your own health at present?" and "What do you think of your own health condition compared to that of other men=women of your age?". Using a five-point Likert scale, respondents were asked to answer from "bad" (1) to "excellent (5)." A high score indicated a better perception of health.

Quality Control

In order to control the quality of questionnaire collection, each IP address limit is set in the background to answer once; In order to improve the authenticity of the questionnaire, the questionnaire does not involve personal information, and there is no reward, highlighting the initiative of the respondents; In order to control the validity of the questionnaire, each question item is set as a required item. If there are unanswered questions, they cannot be submitted, so as to avoid the existence of missing values.

Statistical Method

SPSS 23.0 was used for descriptive statistics and correlation analysis, and the PROCESS 3.2 program developed by Hayes was used for moderated mediation test. The bootstrapping sampling times were 5000 and the test level was high α=0.05.

RESULTS AND ANALYSIS

Common Method Deviation Detection

Harman single factor test was used for common method deviation test (28). The results show that there are 6 factors with characteristic roots >1. The variation explained by the first factor is 26.63%, which is less than the critical standard of 40%, indicating that there is no serious common method deviation variation in this study (29).

Comparison of Negative Emotion Levels in Different Demographic Characteristics

There are significant differences in the scores of negative emotions among the elderly with different gender, educational background, whether they have medical insurance or not and whether they are exposed to suspected cases, but there is no significant difference in age and whether they have religious beliefs, as shown in **Table 2**. In terms of gender, the male elderly were significantly higher than the female elderly ($P < 0.001$); In terms of age, there was no significant difference in negative emotion scores between 55 and 65 years old group and > 65 years old group; In terms of educational background, the negative emotion of the elderly with primary school education and below is the highest, and the post test (LSD) shows that it is significantly higher than that with high school education; In terms of medical insurance, the elderly without medical insurance have the highest negative emotion score, which is significantly higher than those with partial medical insurance and complete medical insurance; In terms of whether to contact suspected cases, the negative emotions of the elderly with suspected cases were significantly higher than those without contact ($P < 0.01$).

Descriptive Statistical Analysis and Correlation Analysis

The mean value, standard deviation and correlation matrix of each variable are shown in **Table 3**. Epidemic concern was significantly positively correlated with negative emotion and negatively correlated with subjective economic status; Negative emotion was negatively correlated with subjective economic status and physical health perception; There was no significant correlation between physical health perception and epidemic

TABLE 1 | Diagnostic criteria for DASS-21 scores.

	Depression	Anxiety	Stress
None	0–4	0–3	0–7
Mild	5–6	4–5	8–9
Medium	7–10	6–7	10–12
heavier	11–13	8–9	13–16
Serious	≥14	≥10	≥17

TABLE 2 | Comparison of negative emotion scores of the elderly with different demographic characteristics.

Project	Number of people	Negative emotion	*T/F*	*P*
Male	179	52.48 ± 13.68	4.897	<0.001
Female	139	44.79 ± 14.15		
55–65 age	201	50.20 ± 14.72	1.768	0.078
>65 age	117	47.26 ± 13.66		
Primary school education and below	28	56.14 ± 14.50	3.017	0.030
Junior high school education	28	49.11 ± 13.08		
High school education (including technical secondary school)	217	47.88 ± 14.53		
University degree or above	45	50.76 ± 13.36		
No religion	288	49.44 ± 14.39	1.236	0.217
Have religion	30	46.03 ± 14.20		
No health insurance	6	69.00 ± 5.87	6.983	0.001
Some medical insurance	102	50.29 ± 14.51		
Fully medicare	210	47.98 ± 14.07		
Contact with suspected cases	8	60.75 ± 8.83	3.698	0.006
Non-contact with suspected cases	310	48.82 ± 14.38		

concern and subjective economic status. According to the establishment criteria of intermediary variables and regulatory variables proposed by Wen et al., if a variable has little correlation with independent variables or dependent variables, it cannot become an intermediary variable, but it can become a better regulatory variable (30). It can be seen that subjective economic status is significantly negatively correlated with epidemic concern and negative emotion, which can be used as an intermediary variable; However, there is no significant correlation between physical health perception and epidemic concern, which is a better regulatory variable.

Moderated Mediation Model Test

According to Wen et al. (31, 32), a moderated mediation model needs to meet the following four conditions: (1) The effect of independent variables on dependent variables is significant; (2) The effect of independent variables on mediator variable is significant; (3) The effect of intermediary variables on dependent variables is significant; (4) The interaction between independent variables and moderator variable has significant effect on mediator variable, or the interaction between mediator variable and moderator variable has significant effect on dependent

TABLE 3 | Correlation coefficients between variables.

	M	*SD*	1	2	3	4
Epidemic concern	4.62	0.52	—			
Negative emotions	49.12	14.38	0.492**	—		
Subjective economic status	3.71	0.85	−0.138*	−0.455**	—	
Perception of physical health	15.81	3.45	−0.047	−0.277**	−0.033	—

*N = 318; *p < 0.05; **p < 0.01; ***p < 0.001.*

variables. Therefore, this article needs to test: (1) epidemic concern has a significant effect on negative emotions; (2) Epidemic concern has a significant effect on subjective economic status; (3) Subjective economic status has a significant effect on negative emotion; (4) The interaction between epidemic concern and physical health perception has a significant effect on subjective economic status, or the interaction between subjective economic status and physical health perception has a significant effect on negative emotion. The first three conditions are the test of the mediating effect of subjective economic status between epidemic concern and negative emotion, and the fourth condition is the test of the moderating effect of physical health perception on this mediating model. Before the test, according to the standardization requirements of mediation and moderating effects, this study centralizes all variables, and then multiplies the scores of physical health perception and epidemic concern, as well as the scores of physical health perception and subjective economic status as interactive items (33).

(1) Intermediary model test of subjective economic status

The first three conditions were tested by multiple stepwise regression analysis, that is, to verify the intermediary role of subjective economic status between epidemic concern and negative emotion. The results showed that in the first step regression equation, epidemic concern had a significant positive predictive effect on the negative emotions of the elderly ($\beta = 0.492$, $t = 10.058$, $P < 0.001$), indicating that epidemic concern can promote negative emotion; In the second step regression equation, epidemic concern has a significant negative predictive effect on subjective economic status ($\beta = -0.138$, $t = -2.479$, $P < 0.05$); The third step is to incorporate the independent variable epidemic concern and the intermediary variable subjective economic status into the regression equation. It is found that the subjective economic status has a significant negative predictive effect on the negative emotion of the elderly ($\beta = -0.395$, $t = -8.917$, $P < 0.001$), indicating that subjective economic status can delay negative emotions, and emotional information attention can still significantly predict negative emotions ($\beta = 0.438$, $t = 9.899$, $P < 0.001$), indicating that subjective economic status plays a partial mediating role between epidemic concern and negative emotion. As shown in **Table 4**, the first three conditions have been met.

(2) The moderator role of physical health perception in mediating model

TABLE 4 | Intermediary model test of subjective economic status.

Regression equation		Overall fitting index				Significance of regression coefficient	
Result variable	**Predictive variable**	***R***	***R*²**	***SE***	***F***	***β***	***t***
Negative emotion	Epidemic concern	0.492	0.242	0.872	101.160***	0.492	10.058***
Subjective economic status	Epidemic concern	0.138	0.019	0.992	6.146*	−0.138	−2.479*
Negative emotion	Epidemic concern	0.629	0.395	0.780	102.900***	0.438	9.899***
	Subjective economic status					−0.395	−8.917***

*N = 318; *p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 5 | Moderated mediation model test.

Result variable	Predictive variable	*R*	*R*²	*SE*	*F*	β	*t*	95% CI
Subjective economic status		0.381	0.145	0.863	17.725***			
	Epidemic concern					−0.154	−2.954**	[−0.257 to −0.052]
	Physical health perception					−0.103	−1.944	[−0.208∼0.001]
	Epidemic concern × Physical health perception					0.403	6.752***	[0.286∼0.521]
Negative emotion		0.692	0.479	0.530	57.323***			
	Epidemic concern					0.416	9.585***	[0.331∼0.502]
	Subjective economic status					−0.369	−8.350***	[−0.456 to −0.282]
	Physical health perception					−0.248	−5.914***	[−0.330 to −0.165]
	Epidemic concern × physical health perception					−0.099	−1.946	[−0.199∼0.001]
	Subjective economic status × physical health perception					0.055	1.287	[−0.029∼0.139]

*N = 318; *p < 0.05; **p < 0.01; ***p < 0.001.*

The fourth condition uses model 59 in the process program developed by Hayes to test the moderated mediation model, and takes physical health perception as the regulating variable into the equation. The results are shown in **Table 5**. Epidemic concern and physical health perception have significant negative prediction on subjective economic status, and their interaction items have significant positive prediction on subjective economic status ($\beta = 0.403$, $P < 0.001$), indicating that physical health perception has a moderating effect on epidemic concern → subjective economic status; Epidemic concern and physical health perception have significant negative predictive effects on negative emotions, but the interaction between epidemic concern and physical health perception and the interaction between subjective economic status and physical health perception have no significant predictive effects on negative emotions ($\beta_{\text{Interactionitem 1}} = -0.099$, $P > 0.05$; $\beta_{\text{Interactionitem2}} = 0.055$, $P > 0.05$), indicating that physical health perception has no moderating effect on epidemic concern → negative emotion, subjective economic status → negative emotion. It can be seen that physical health perception plays a moderator role in the first half path of epidemic concern → subjective economic status → negative emotion mediation model, but does not play a regulatory role in the second half path of mediation model and in the direct path of epidemic concern to negative emotion.

In order to specifically explore the moderating effect of physical health perception, add or subtract a standard deviation from the average value of physical health perception and epidemic concern score, and divide it into low physical health perception group, high physical health perception group, low epidemic concern group and high epidemic concern group. Make the slope diagram of moderating effect (as shown in **Figure 2**) to further analyze its simple effect. The results showed that in the group with low level of physical health perception, epidemic concern significantly negatively predicted subjective economic status ($t = -6.888$, $P < 0.001$); In the group with high level of physical health perception, epidemic concern significantly positively predicted subjective economic status ($t = 3.201$, $P < 0.01$). The results show that physical health perception plays a regulatory role between epidemic concern and subjective economic status.

DISCUSSION

This study found of the elderly in the COVID-19 period were very serious, the number of elderly people without depression, anxiety or stress in the COVID-19 is 0, and most of them are in serious and severe degrees, with anxiety being the most serious, followed by stress and depression, shown in **Table 6**. 44.03% of the elderly were in a state of serious depression, 94.34% of the elderly were in a state of serious anxiety, and 70.13% of the elderly were in a state of serious stress. It can be seen that the

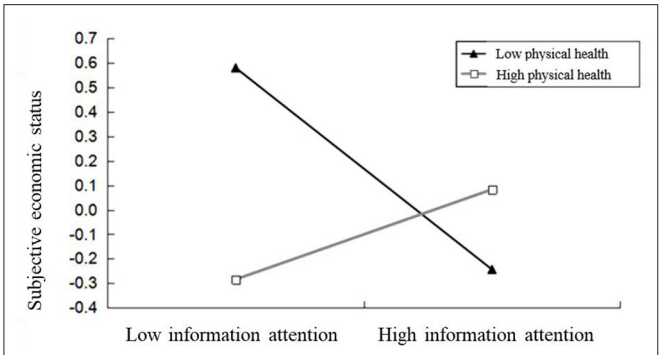


FIGURE 2 | The moderator role of physical health perception between epidemic concern and subjective economic status.

TABLE 6 | Number and proportion of people with different symptoms of depression, anxiety and stress.

	Depression	Anxiety	Stress
None	0	0	0
Mild	40 (12.58%)	0	4 (1.26%)
Medium	61 (19.18%)	4 (1.26%)	40 (12.58%)
heavier	77 (24.21%)	14 (4.40%)	51 (16.04%)
Serious	140 (44.03%)	300 (94.34%)	223 (70.13%)

elderly in the epidemic period, as a vulnerable group, are a group that needs to be cared for in social psychological counseling and crisis intervention.

Like the conclusions of other studies, the anxiety, depression and stress of the elderly men are significantly higher than those of women (2). The elderly in different age groups do not score significantly in negative emotions, because COVID-19 is a high-intensity stressful life event, which poses a major threat to the mental health of all people, and has become an important stressor causing people's psychological balance disorder (34). The elderly with primary school education and below have the highest negative emotions, which may be that the teachers with low education level lack effective information about the epidemic and lack confidence in how to prevent the epidemic, resulting in a large degree of anxiety, stress and depression. Previous studies have found that religious belief has a significant positive impact on the health status of the elderly, and its important way is to participate in religious activities. This social capital helps to alleviate the loneliness of elderly believers (35). This study found that religious beliefs have no significant impact on the negative emotions of the elderly under the epidemic. The reason for this phenomenon is that for the sake of epidemic prevention and control, gathering activities are limited, and religious worship and other activities are also greatly affected. To some extent, this finding shows that the mental health of Chinese elderly believers mainly benefits from external religious orientation rather than internal religious orientation. Different from the research of Zhou Qin and others, old-age insurance such as the new rural cooperative medical insurance, as a stable source of income, has played a positive role in improving residents' mental health, while the corresponding role of medical insurance systems such as the new rural cooperative medical system is basically not significant (36). This study found that medical insurance can greatly improve the mental health of the elderly, which is related to the special period of the epidemic. The elderly exposed to suspected cases have significant negative emotions because they are worried about being infected and are transmitted to their families.

Correlation analysis showed that epidemic concern was significantly positively correlated with negative emotion and negatively correlated with subjective economic status; Negative emotions are significantly negatively correlated with subjective economic status and physical health perception. Further analysis shows that epidemic concern has a significant positive predictive effect on negative emotions. This conclusion is consistent with the previous research results (3). The more concerned about epidemic information, the more sensitive to various information, which is more likely to lead to certain negative emotions.

The mediating effect model shows that subjective economic status plays a partial mediating role between epidemic concern and negative emotion, indicating that epidemic concern directly leads to the generation of negative emotional experience of the elderly on the one hand, and indirectly reduces the generation of negative emotional experience of the elderly through subjective economic status on the other hand. This shows that economic status is an important factor affecting the physical and mental health of the elderly in Normalization Period of COVID-19. Existing studies have also confirmed that Socio economic status is one of the fundamental factors affecting people's physical and mental health (37); the improvement of economic level will help individuals obtain better positive psychological resources, which will directly improve the mental health of the elderly and reduce the occurrence of emotional and behavioral problems (38).

The interaction between epidemic concern and physical health perception has a significant positive prediction on subjective economic status, indicating that physical health perception plays a moderating effect in the impact of epidemic concern on subjective economic status. Further simple slope analysis shows that different levels of physical health perception play different roles in the relationship between epidemic concern and subjective economic status: the older people with low-level physical health perception pay more attention to the epidemic, the lower their subjective economic status, and the lower their attention to the epidemic, the higher their subjective economic status; Different from the elderly with high-level physical health perception, the lower their attention to the epidemic, the lower their subjective economic status. The more they pay attention to the epidemic, the higher their subjective economic status. According to the theory of limited cognitive resources, people's cognitive resources are limited, which makes people have to allocate limited resources to different activities or different aspects of the same activity (39). When the elderly are weak and ill, they will continue to pay attention to their health, so that their cognitive resources have been occupied. Therefore, those who allocate more cognitive resources to obtain epidemic information are often those with poor economic conditions.

Compared with those with good economic conditions, they are more likely to perceive the threat of the epidemic to their health; Correspondingly, the cognitive resources of healthy elderly people do not focus too much on their own physical health. They pay more attention to the external world. The more types of activities they participate in, the better their health will benefit. These are closely related to economic conditions (40). Therefore, those who pay more attention to epidemic information are often those with better economic conditions.

CONCLUSION

To sum up, this study found that subjective economic status and physical health perception play a regulatory intermediary role between epidemic concern and negative emotions, which is of great significance to prevent and reduce the occurrence of psychological crisis and effectively carry out psychological counseling for the elderly during the epidemic period. Firstly, the intermediary role of subjective economic status suggests that we can start from the aspects of medical insurance, ensuring the living standard of low-income people and increasing income; Secondly, the differential moderating effect of physical health perception suggests that different targeted strategies should be adopted in prevention and intervention. For the elderly with low physical health, we should focus on life security and physical health maintenance, while for the elderly with good physical health, we should release effective epidemic information to guide them to actively participate in epidemic prevention and control.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

KX: research, data analysis, writing, and revising papers.

FUNDING

This work was supported by Henan Key R&D and Promotion Special Project (soft science research) (212400410306); Henan Philosophy and Social Science Planning Project (2021BZZ007); National Natural Science Foundation of China (72074235); Henan Research and Practice Project of Higher Education Teaching Reform (2021SJGLX496).

REFERENCES

1. Liu HL, Tian QH. Ways to improve the mental health level of the elderly from the perspective of community activities. *Chin J Gerontol.* (2019) 14:3571–6. doi: 10.3969/j.issn.1005-9202.2019.14.067
2. Chen XY, Chen Z, Ma JL. Study on the psychological status and influencing factors in the elderly living in nursing home under long-term fully closed management during the epidemic period of COVID-19 in low risk area. *Practical Geriatrics.* (2020) 12:1329–32. doi: 10.3969/j.issn.1003-9198.2020.12.033
3. Gao R, Zhang BZ, Peng SR. Mental health status of the general population during the COVID-19 epidemic in Ya'an City. *Chin Mental Health J.* (2021) 1:79–84. doi: 10.3969/j.issn.1000-6729.2021.01.014
4. Wang YL, Chen H, Chen HJ. Research on the status quo and influencing factors of frailty among the elderly in the community. *Chin Clin Nurs.* (2022) 4:203–6. doi: 10.3969/j.issn.1674-3768.2022.04.002
5. Yang ZK, Shi XN, Yang XT. Discussion on the health management mode of the elderly with chronic diseases under the background of the normalization of epidemic prevention and control. *Chin J Public Health Manage.* (2021) 5:616–9. doi: 10.19568/j.cnki.23-1318.2021.05.0013
6. Hardy GE, Woods D, Wall TD. The impact of psychological distress on absence from work. *J Appl Psychol.* (2003) 2:306–14. doi: 10.1037/0021-9010.88.2.306
7. Wang JX, Ying XP. Cognition, emotion and action: social mentality in the emergency response. *Exploration Free Views.* (2020) 4:232–43.
8. Brashers DE. Communication and uncertainty management. *J Commun.* (2001) 3:477–97. doi: 10.1111/j.1460-2466.2001.tb02892.x
9. Sprengelmeyer R, Jentzsch I. Event related potentials and the perception of intensity in facial expressions. *Neuropsychologia.* (2006) 14:2899–906. doi: 10.1016/j.neuropsychologia.2006.06.020
10. Maeland JG. Socioeconomic inequalities in health. *Norsk Epidemiologi.* (2009) 1:557–66. doi: 10.5324/nje.v12i1.498
11. Jiang GR, Li DY, Ren ZH, Yupeng Y, Xinchun W, Xu Z, et al. The status quo and characteristics of Chinese mental health literacy. *Acta Psychol Sinica.* (2021) 2:182–201. doi: 10.3724/SP.J.1041.2021.00182
12. Wang XH. The impact of socio-economic status on the subjective well-bei g of the elderly. *J Dalian University Technol.* (2021) 3:92–100. doi: 10.19525/j.issn1008-407x.2021.03.011
13. Pu C, Tang GJ, Huang N, Chou YQ. Predictive power of self-rated health for subsequent mortality risk during old age: analysis of data from a nationally representative survey of elderly adults in Taiwan. *J Epidemiol.* (2011) 4:278–84. doi: 10.2188/jea.JE20100131
14. Pu CC, Wang HL. Psychological reactions and coping strategies of the elderly facing COVID-19. *Chin J Mental Health.* (2020) 3:257–8. doi: 10.3969/j.issn.1000-6729.2020.3.024
15. Pressman SD, Cohen S. Does positive affect influence health? *Psychol Bull.* (2005) 6:925. doi: 10.1037/0033-2909.131.6.925
16. Lovibond P F, Lovibond S H. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther.* (1995) 3:335–43. doi: 10.1016/0005-7967(94)00075-U
17. Li XL, Tang HB, Guo F. A review of depression-anxiety-stress scale's reliability and validity. *Chin J Clin Psychol.* (2012) 3:350–2. doi: 10.16128/j.cnki.1005-3611.2012.03.039
18. Gong X, Xie XY, Xu R, Luo Y. Psychometric properties of the Chinese versions of DASS-21 in Chinese college students. *Chin J Clin Psychol.* (2010) 4:443–6.
19. Wen Y, Wu DX, Lv XJ, Li H. Psychometric properties of the Chinese short version of depression anxiety and stress scale in Chinese adults. *Chin J Public Health.* (2012) 11:1436–8. doi: 10.11847/zgggws2012-28-11-14
20. Sun Y, Yan F, Wang WT. Study on the moderated effect of migration years on the relationship between stress and depression and anxiety among the elderly in new cities. *General Pract China.* (2017) 2:210–3. doi: 10.3969/j.issn.1007-9572.2017.02.018
21. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *Br J Clin Psychol.* (2003) 2:111–31. doi: 10.1348/014466503321903544
22. Lovibond SH, Lovibond PF. *Manual for the Depression anxiety Stress Scales*. Sydney, NSW: Psychology Foundation. (1995). doi: 10.1037/t01004-000
23. Wang JX, Gao WJ, Chen MQ. Investigation report on social psychology under the COVID-19. *Governance.* (2020) 1:55–64. doi: 10.16619/j.cnki.cn10-1264/d.2020.z1.014

24. Adler NE, Epel ES, Castellazzo G, Ickovics JR. Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychol.* (2000) 6:586–92. doi: 10.1037/0278-6133.19.6.586
25. Shang SJ, Bai BY, Zhong N. Family social class and meaning in life: mediating of basic psychological need satisfaction. *Chin J Clin Psychol.* (2016) 6:1108–11. doi: 10.16128/j.cnki.1005-3611.2016.06.032
26. Heo J, Lee Y. Serious leisure, health perception, dispositional optimism, and life satisfaction among senior games participants. *Educ Gerontol.* (2010) 2:112–26. doi: 10.1080/03601270903058523
27. Liu F. *Relationship Between Health Perception and Death Anxiety and Its Mechanism of Old People*. Sichuan: Sichuan Normal University. (2015).
28. Zhou H, Long LR. Statistical remedies for common method biases. *Adv Psychol Sci.* (2004) 6:942–50.
29. Xiong HX, Zhang J, Ye BJ. Common method variance effects and the models of statistical approaches for controlling it. *Adv Psychol Sci.* (2012) 5:757–69. doi: 10.3724/SP.J.1042.2012.00757
30. Wen ZL, Hou JT, Zhang L. A comparison of moderator and mediator and their applications. *Acta Psychol Sinica.* (2005) 2:268–74. doi: 10.1111/j.1744-7909.2005.00136.x
31. Wen ZL, Zhang L, Hou JT. Mediated moderator and moderated mediator. *Acta Psychol Sinica.* (2006) 3:448–52. doi: 10.1016/S0379-4172(06)60092-9
32. Wen ZL, Liu HY, Hou JT. *Analysis of Moderating and Mediating Effect*. Beijing: Educational Science Press. (2012) 91–2.
33. Wen ZL, Hou JT, Marsh HW. Appropriate standardized estimates for moderating effects in structural equation models. *Acta Psychol Sinica.* (2008) 6:729–36. doi: 10.3724/SP.J.1041.2008.00729
34. Huang C, Wang Y, Wang Y, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel corona virus in Wuhan, China. *Lancet.* (2020) 10:497–506. doi: 10.1016/S0140-6736(20)30183-5
35. Jiang QC, Zhang KZ. Does religious belief affect the health of the elderly? *World Econ Papers.* (2013) 5:85–106.
36. Zhou Q, Jiang WG, Guo X. The effect of social insurance on mental health among rural residence: an empirical analysis based on CHARLS data. *China Econ Stud.* (2018) 5:125–36. doi: 10.19365/j.issn1000-4181.2018.05.10
37. Hu JR. The relationship between socio-economic status and physical function and health of the elderly – taking exercise and medical insurance as intermediary variables. *J Beijing University Sci Technol.* (2019) 4:61–9. doi: 10.3969/j.issn.1008-2689.2019.04.009
38. Lei XY, Sun XT, Strauss J, Zhang P, Zhao Y. Depressive symptoms and SES among the mid-aged and elderly in China: Evidence from the China Health and Retirement Longitudinal Study national baseline. *Soc Sci Med.* (2014) 11:224–32. doi: 10.1016/j.socscimed.2014.09.028
39. Kahneman D, Treisman A. Changing views of attention and automaticity. In: R Parasuraman, DR Davis. *Varieties of Attention*. Orlando: Academic Press (1984) 29–61.
40. Ding ZH. The influence of social participation of the rural oldest-old on their health. *Lanzhou Acad J.* (2018) 12:179–95. doi: 10.3969/j.issn.1005-3492.2018.12.019

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Oncology

METHODS
published: 29 June 2022
doi: 10.3389/fonc.2022.932496

Performance Evaluation of the Deep Learning Based Convolutional Neural Network Approach for the Recognition of Chest X-Ray Images

Sandhya Sharma[1], Sheifali Gupta[2], Deepali Gupta[2], Junaid Rashid[3], Sapna Juneja[4], Jungeun Kim[3,5]* and Mahmoud M. Elarabawy[6,7]*

[1] Chitkara University Institute of Engineering and Technology, Chitkara University, Baddi, India, [2] Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, India, [3] Department of Computer Science and Engineering, Kongju National University, Cheonan, South Korea, [4] KIET Group of Institutions, Ghaziabad, India, [5] Department of Software, Kongju National University, Cheonan, South Korea, [6] Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia, [7] Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia, Egypt

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu, VIT University, India

Reviewed by:
Sumit Badotra, Lovely Professional University, India
Sharma Tripti, Maharaja Surajmal Institute of Technology, India

***Correspondence:**
Junaid Rashid
junaidrashid062@gmail.com
Jungeun Kim
jekim@kongju.ac.kr

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 April 2022*
Accepted: *01 June 2022*
Published: *29 June 2022*

Citation:
Sharma S, Gupta S, Gupta D, Rashid J, Juneja S, Kim J and Elarabawy MM (2022) Performance Evaluation of the Deep Learning Based Convolutional Neural Network Approach for the Recognition of Chest X-Ray Images. Front. Oncol. 12:932496. doi: 10.3389/fonc.2022.932496

Recent advancement in the field of deep learning has provided promising performance for the analysis of medical images. Every year, pneumonia is the leading cause for death of various children under the age of 5 years. Chest X-rays are the first technique that is used for the detection of pneumonia. Various deep learning and computer vision techniques can be used to determine the virus which causes pneumonia using Chest X-ray images. These days, it is possible to use Convolutional Neural Networks (CNN) for the classification and analysis of images due to the availability of a large number of datasets. In this work, a CNN model is implemented for the recognition of Chest X-ray images for the detection of Pneumonia. The model is trained on a publicly available Chest X-ray images dataset having two classes: Normal chest X-ray images and Pneumonic Chest X-ray images, where each class has 5000 Samples. 80% of the collected data is used for the purpose to train the model, and the rest for testing the model. The model is trained and validated using two optimizers: Adam and RMSprop. The maximum recognition accuracy of 98% is obtained on the validation dataset. The obtained results are further compared with the results obtained by other researchers for the recognition of biomedical images.

Keywords: biomedical images, convolutional neural network, deep learning, chest X-rays, optimizers

1 INTRODUCTION

Fungi, bacteria, or viruses fill the lungs with fluid, which is the main reason for the infection of the lungs. This infection which is known as Pneumonia causes effusion of the pleural as well as inflammation in the air sacs. Pneumonia is the cause of 15% of the deaths in children who are below the age of 5 years (1). Pneumonia is more prevalent in underdeveloped or developing 'countries with scanty medical resources. This disease can be fatal if not diagnosed and treated early. It can cause the failure of the respiratory system or inflammation can be uncontrolled. For its diagnosis: Magnetic

resonance imaging (MRI), radiography (X-ray), or computed tomography (CT) is frequently used, whereas radiography or X-ray is most commonly used as it is an inexpensive and non-invasive examination. Chest X-ray is the radiology test in which the chest is exposed to radiation to produce an image of the internal organs of the chest. Once the chest X-ray is produced then the radiologist interprets the organs of the chest by looking at the image produced. An automated system is required which can detect the Pneumonia for the X-ray image as manual chest X-ray examination is prone to variability (2). Sometimes the X- rays are not very clear, which makes it difficult for the radiotherapist to identify the features that help in the detection of disease. Therefore, a deep learning-based CNN model is required that can automatically recognize the healthy and pneumonic lungs from the Chest X-ray images (3). Deep learning and machine learning in the field of Artificial Intelligence that is being used in various fields like visual recognition, natural language processing (NLP), self-driving cars, fraud detection, news aggregation, and so on. Deep Learning based Models work similar to that of a human brain, by using data as the input so that it can be processed through various layers for its recognition. These models can find the hidden features that are sometimes not possible to detect by radiologists. CNN is the main tool of deep learning models which can automatically find the various features of the images which makes the recognition of images possible for different classes.

Figure 1 shows the X-ray images of the lungs for the healthy and Pneumonic chest. **Figure 1A** is the healthy X-ray whereas **Figure 1B** is the Pneumonic X-ray. It can be seen that the pneumonic X-ray has white spots called infiltrates.

The paper is structured as: A review similar to the related work is given in Section 2; while Section 3 represents the methods used in this paper. The results obtained are depicted in Section 4 which is followed by the conclusion in Section 5. The contributions of the proposed work are given below:

- A CNN based model is developed which can be used for the recognition of chest X-ray images of the lungs.
- A model is evaluated to find the suitable optimizer, number of epochs as well as the learning rate (LR) which can work well on the available dataset.
- A Comparison of the proposed work with other developed models for the recognition of biomedical images.

2 RELATED WORK

Detect Pneumonia using chest X-ray images is a problem for many years (4, 5). Sometimes the chest X-rays are not clear enough to identify the 'features' that help in the diagnosis of the disease. Various machine learning, as well as deep learning models, are being used for the detection and classification of pneumonia. In machine learning techniques, various features are required to be manually extracted whereas features are automatically extracted by employing deep learning techniques for detection and classification. In this section, an attempt has been made to present the work that has been done for the recognition of biomedical images using Machine Learning as well as Deep Learning techniques. Chandra et al. (6) have employed a machine learning technique to detect pneumonia. First, the lung regions are segmented from the X-ray, and later eight statistical features are extracted from the regions which are used for classification. Once the features are extracted, various classifiers are implemented for classification like the random forest, logistic regression, and multilayer perceptron (MLP) technique. The model is evaluated on 412 images and the accuracy obtained is 95.39%. Yue et al. (7) have extracted six features from the 52 CT scanned images for the detection of pneumonia. An AUC (area under the curve) of 97% is achieved. Similarly, Kuo et al. (8) have also manually extracted various features for the detection and classification of Pneumonia. The extracted features are applied to various classification models such as decision tree, support vector machine (SVM), and logistic regression. The maximum accuracy obtained is 94.5% using the decision tree.

Features are manually extracted in machine learning techniques (9, 10) whereas features are automatically extracted in deep learning techniques (11, 12). Tuncer et al. (13) has also employed a machine learning-based technique, which has further imposed the fuzzy tree transformed technique on

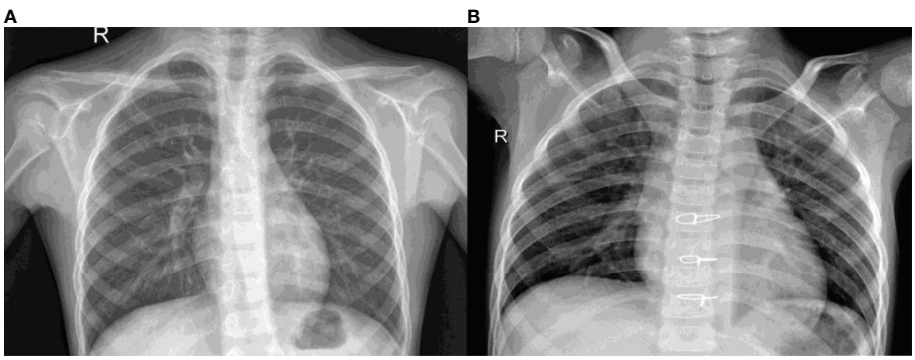


FIGURE 1 | X-ray images of the lungs **(A)** Healthy chest X-ray **(B)** Pneumonic chest X-ray.

the images. Then, various features are extracted with the help of multi-kernel local binary pattern, and later various classifiers are employed. Obtained accuracy is 97.01%. Various deep learning techniques are being employed for the detection and classification of chest X-rays.

Sharma et al. (14) implemented a CNN based model for the classification of pneumonia in Chest X-rays. The data augmentation technique is also employed to increase the data. Obtained accuracy is 90.68%. Similarly, Stephen et al. (15) also employed CNN based model by incorporating a data augmentation technique and obtained an accuracy of 93.73%. Rajpukar et al. (16) have also implemented a Deep learning-based DenseNet-121 CNN model for classifying pneumonia from Chest X-ray images. F1 score of only 76.8% is achieved. The model has not perform well due to the unavailability of the patient's history. It has been found that various machine learning and deep learning models have been implemented for recognition. Manual extraction of features is a cumbersome task as required in machine learning models, but features are not required to be manually extracted in deep learning models, which makes them easier to implement. In this work, we have obtained a recognition accuracy of 98% using a 12 layered customized CNN model with the help of RMSprop Optimizer. A suitable optimizer has also been obtained for the available dataset. The prepared model has also been evaluated using Adam Optimizer while other parameters are kept the same.

3 PROPOSED APPROACH

CNN based model is proposed for the recognition and classification of the Chest X-ray images. Publically available Kaggle dataset of X-ray images is used for recognition. The proposed Methodology of the work is given in **Figure 2** below. It can be seen from the figure that an X-ray image is given as input to the CNN model. First, the image is resized to 64X32 and then it will pass through some different layers (convolution, pooling, and flattening) after dividing into training and validation datasets. In the end, the obtained image will be passed through the output class which will predict the class of that image i.e. Pneumonic or Healthy image.

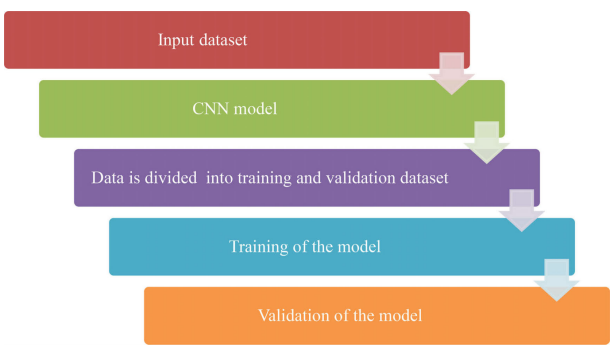


FIGURE 2 | Methodology of the proposed work.

3.1 Input Dataset

Kaggle dataset is used for classification and recognition. Data preprocessing is not required as the dataset is already available in a clean and pre-processed format. The dataset consists of two classes of chest X-ray images where each class has 5000 images of Healthy and Pneumonic Chest X-rays. 4000 images of each class are used to train the model, and 1000 images of each class are used to test the model. Few of the dataset images for Healthy and Pneumonic Chest X-rays can be observed in **Figure 3**.

3.2 Data Augmentation

To increase the number of images for each class, a data augmentation technique is employed. As data augmentation helps in improving the performance of the model a model is trained on different examples of the dataset. Flipping and Rotation of the existing dataset are carried out to enhance the dataset. **Figure 4** is showing the flipped and rotated images of the dataset.

3.3 Proposed Convolution Model

A total of 10000 images of healthy and pneumonic chest X-rays have been taken and each class has 5000 samples.

All the samples are passed through the set of three main layers of the CNN model: convolution, pooling and then the fully connected layer. The basic block diagram of the proposed model is shown in **Figure 5**. CNN model has been the biggest contributor in the field of image analysis and computer vision. CNN model or ConvNet can learn the features of various images which are used for their classification and recognition. The convolution layer helps in obtaining the feature maps from the images which are the extraction of various features. Automatic feature extraction takes place using a convolution layer. The pooling layer helps in retaining the important features while it also helps in dropping off the features which are least important for the recognition. This layer helps reduce the image size. The fully connected layer at the end helps in obtaining the output as classes. The architecture of the CNN model is shown in **Figure 6**. The X-ray image is given as input to the first layer of the model, i.e. convolution layer with 32 filters and the size of each filter is 3x3 which will help in obtaining 32 feature maps. These are given to the next layer: the Maxpooling layer with a filter pool of size 2x2 and a stride with the size (2, 2). All Maxpooling layers used in the model have the same size. In the next convolution layer, the number of filters is increased from 32 to 64 with the same size 3x3 as used earlier, which is again followed by a max-pooling layer. The number of filters in the next convolution layer increases to 128 which is followed by the max-pooling layer, and the last convolution layer has 256 filters (17, 18). In the end, a flattening layer is introduced so that the obtained pooled feature maps can be fed to the array of the neural network (19, 20). Two dense layers are introduced after the flattening layer which is followed by an output layer.

4 RESULTS AND ANALYSIS

In this section, results obtained for the recognition of chest X-rays are shown. The CNN model is trained using 80% of the data for the recognition of healthy and pneumonic chest X-rays. Once the model is trained, then it is validated using the rest of the

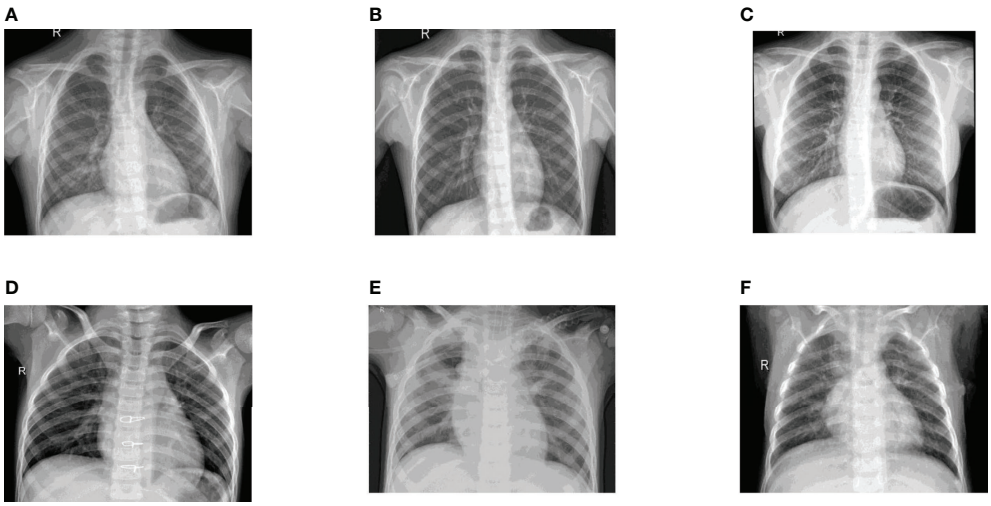


FIGURE 3 | Dataset Images: **(A)** Healthy Chest X-ray 1, **(B)** Healthy Chest X-ray 2, **(C)** Healthy Chest X-ray 3, **(D)** Pneumonic Chest X-ray 1, **(E)** Pneumonic Chest X-ray 2 and **(F)** Pneumonic Chest X-ray 3.

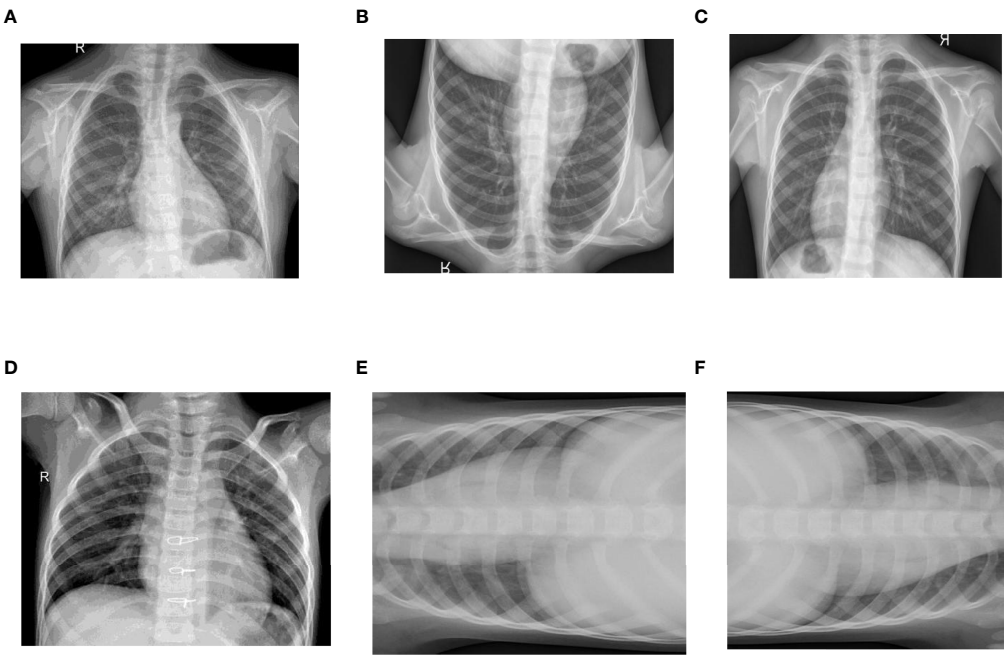


FIGURE 4 | Data Augmentation of Images: **(A)** Original Healthy Chest X-ray, **(B)** Healthy Chest X-ray Flipped Vertically, **(C)** Healthy Chest X-ray Flipped Horizontally, **(D)** Original Pneumonic Chest X-ray, **(E)** Pneumonic Chest X-ray rotated 90 degrees left and **(F)** Pneumonic Chest X-ray rotated 90 degrees right.

images. The model is trained using RMSprop as well as Adam optimizer with an LR of 0.01 and the number of epochs used is 30. The model has not perform well with the other optimizers as compared to the two selected optimizers. Even, when the number of epochs are increased from 30, the model's performance started to deteriorate or either remained the same. Therefore, the model performed at its best with the selected optimization parameters. All the results are obtained on the Google Colab, which is a free Jupyter environment that runs on the cloud.

4.1 Result Analysis Using RMSprop Optimizer

The model has been trained and validated with the help of RMSprop Optimizer where the number of epochs using which the model is trained is 30, LR used is 0.01 and Batch size of 4 is

FIGURE 5 | Block diagram of the CNN Model.

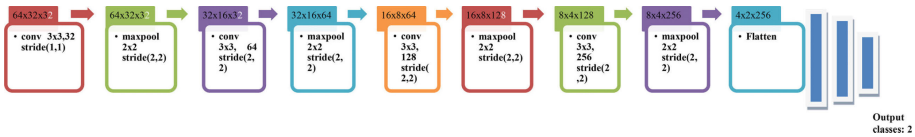


FIGURE 6 | Layered Architecture of the CNN Model.

used. The model is validated with the help of various parameters such as accuracy, loss, recall, area under the curve (AUC) and precision (19, 21).

4.1.1 nTraining and Validation: Accuracy and Loss Curves

Figure 7 is showing the results obtained on the training as well as the validation dataset. A maximum validation accuracy of 98% is obtained on the 29th epoch. In **Figure 7A**, Validation accuracy (22) is below 60% on the 1st epoch and increases till 3rd epoch to around 84% and then decreases on 4th epoch and again increases on the next epoch (23–25). The highest validation accuracy obtained is 98% on the 29th epoch, and it is 97% on the last epoch. Training accuracy also started at around 80% in the 1st epoch and going close to 90% in the last epoch. **Figure 7B** is showing the curve for the training and validation dataset. Validation loss is around 1.61 in the 1st epoch and is decreasing till the 5th epoch to 0.13 but again increasing in the 6th epoch to 1.74. The value obtained for validation loss is 0.06 on the last epoch but the minimum validation loss obtained is 0.05 on the 29th epoch itself. Training loss is 0.50 in the 1st epoch and decreases to close to a value of 0.30 in the last epoch.

4.1.2 Confusion Matrix Parameters Analysis

Performance metrics are used for the recognition of patterns and the classification of images. These parameters are important in building a perfect model which can give more accurate and precise results. Maximum validation precision, maximum validation recall, and maximum validation AUC obtained are 0.05, 98%, 98% and 99% on the 29th epoch.

The best results for the available dataset using the proposed CNN model are obtained on the 29th Epoch. **Figure 8A** shows the curve of precision for the training and validation dataset. The value of precision for the validation data is below 0.60 on the 1st epoch, increasing till the 3rd epoch and later decreasing and then increasing. The maximum value of the precision is 0.98 on the 29th epoch, while 0.97 on the last epoch. The value of precision for the training dataset is around 0.80 on the 1st epoch and goes close to 0.90 on the last epoch. The parameter convergence plot for Recall is shown in **Figure 8B**. Recall for the validation dataset is below 0.60 on the 1st epoch and 0.97 on the last epoch. The maximum value for the recall is 0.98 on the 29th epoch. The value of recall for the training dataset is 0.80 on the 1st epoch and close to 0.90 on the last epoch. The AUC for the validation and training dataset is shown in **Figure 8C**. The highest value of AUC achieved is around 0.99 on the 29th as well as 30th epoch, while the minimum value of AUC is 0.77 on the 2nd epoch and it started at 0.78 on the 1st epoch. The AUC for the training data is around 0.87 on the 1st epoch and is going close to 0.96 on the 30th epoch.

4.2 Result Analysis Using Adam Optimizer

Here the model is trained and validated using Adam Optimizer (19, 20). The number of epochs using which the model is trained is

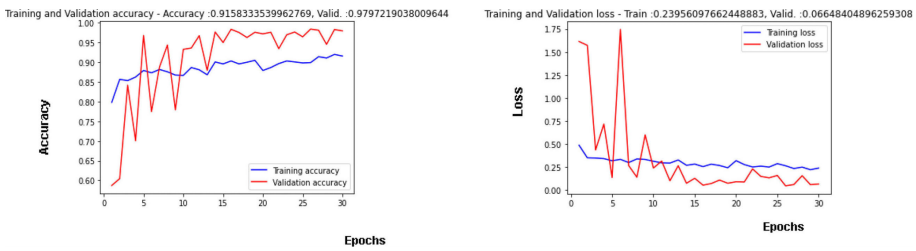


FIGURE 7 | Training and Validation Curves using RMSprop Optimizer: **(A)** Accuracy, **(B)** Loss.

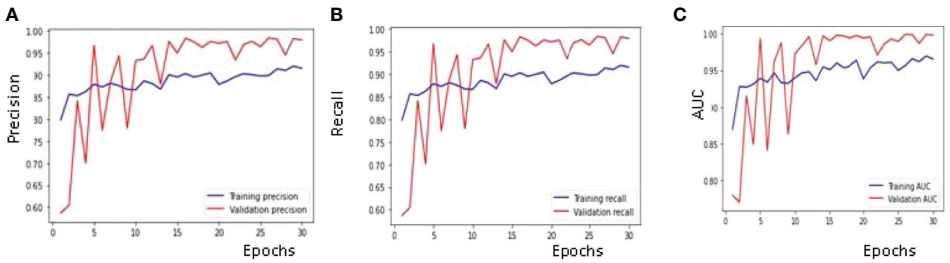


FIGURE 8 | Confusion Matrix Parameter Curves using RMSprop Optimizer: **(A)** Precision, **(B)** Recall and **(C)** AUC.

30, the LR used is 0.01, and a batch size of 4 is used. The model is validated with the help of various parameters like accuracy, loss, Recall, Area under the curve (AUC) (19, 26), and precision (22, 27).

4.2.1 Training and Validation: Accuracy and Loss Curves

Figure 9 is showing the results obtained on the training as well as the Validation dataset. The maximum validation accuracy of 97% is obtained on the 29th epoch. It can be observed from **Figure 9A** that the validation accuracy is below 60% in the 1st epoch and increases till the 4th epoch to around 94% and then decreases in the 5th and 6th epoch and again increases in the next epoch. The highest validation accuracy obtained is 97% on the 29th epoch and it is 94% on the last epoch. Training accuracy also started at around 80% in the 1st epoch and is below 90% till the last epoch. **Figure 9B** is showing the loss curve for the training and validation dataset. The validation loss is around 2.2 in the 1st epoch and is decreasing till the 5th epoch to 0.47 but again increasing in the 6th epoch to 1.25. The value obtained for validation loss is 0.15 on the last epoch but the minimum validation loss obtained is 0.08 on the 29th epoch itself. Training loss is around 2 in the 1st epoch and decreases to close to a value of 0.80 in the last epoch.

4.2.2 Confusion Matrix Parameters Analysis

Maximum validation precision, maximum validation recall, and maximum validation AUC obtained are 98%, 98% and 99% on the 29th epoch. The best results for the available dataset using the proposed CNN model are obtained on the 29th Epoch. **Figure 10A** is shows the curve of precision for the training and validation dataset. The value of precision for the validation data is below 0.60 on the 1st epoch, increasing till the 5th epoch and later decreasing and then increasing. The maximum value of the precision is 0.98 on the 29th epoch, while 0.94 on the last epoch. The value of precision for the training dataset is around 0.80 on the 1st epoch and goes close to 0.87 on the last epoch. The parameter convergence plot for Recall is shown in **Figure 10B**. Recall for the validation dataset is below 0.60 on the 1st epoch and 0.94 on the last epoch. The maximum value for the recall is 0.98 on the 29th epoch. The value of recall for the training dataset is 0.80 on the 1st epoch and close to 0.87 on the last epoch. The AUC for the validation and training dataset is shown in **Figure 10C**. The highest value of AUC achieved is around 0.99 on the 29th and it is 0.98 on the 30th epoch while the minimum value of AUC is 0.67 on the 1st epoch. The AUC for the training data is around 0.87 on the 1st epoch and is going close to 0.93 on the 30th epoch.

4.3 Results Comparison of Adam and RMS Optimizer

The parameter values obtained by using two optimizers: RMSprop and Adam are shown in **Table 1**. The table shows the results for various validation parameters such as accuracy, loss, precision, recall, and AUC (25). A model is trained and validated using 30 epochs, it can be seen that the maximum accuracy is obtained on the 29th epoch using both the optimizers. RMSprop has achieved the maximum accuracy of 98% with a minimum loss of 0.05 while Adam optimizer has obtained the

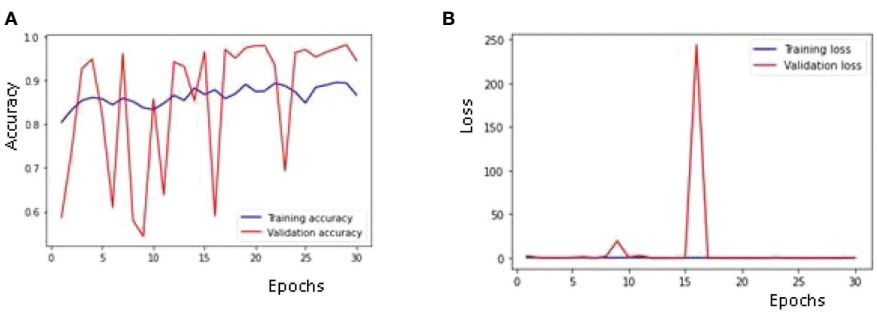


FIGURE 9 | Training and Validation Curves using Adam Optimizer: **(A)** Accuracy, **(B)** Loss.

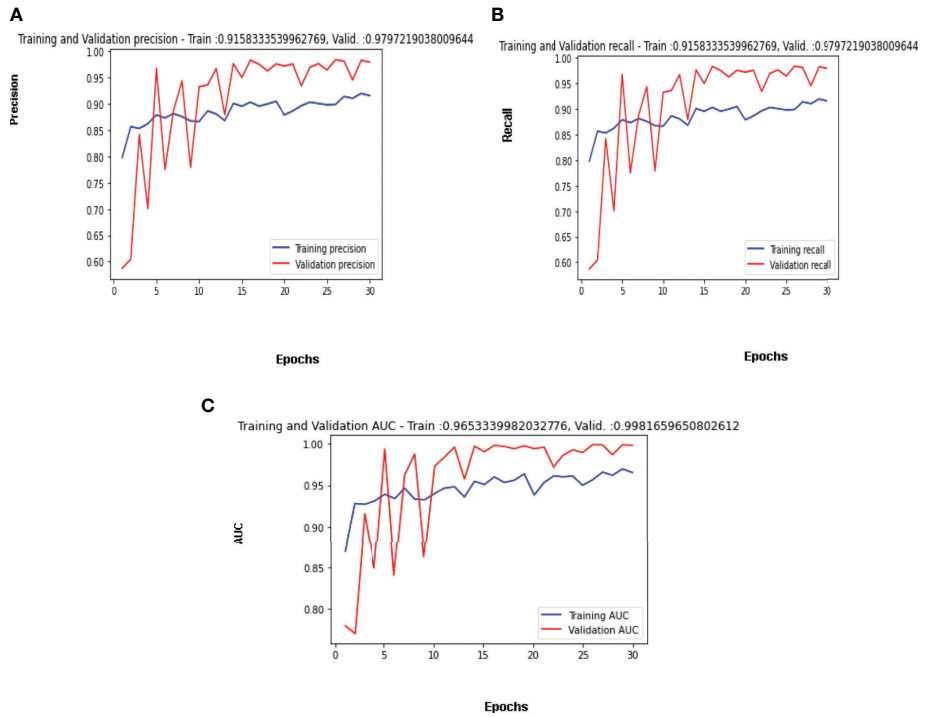


FIGURE 10 | Confusion Matrix Parameter Curves Using Adam Optimizer: **(A)** Precision, **(B)** Recall, and **(C)** AUC.

TABLE 1 | Results obtained on the Validation dataset using two optimizers: RMSprop and Adam.

Optimizer	Epoch	Accuracy	Loss	Precision	Recall	AUC
	1	0.58	1.61	0.58	0.58	0.78
	2	0.60	1.57	0.60	0.60	0.77
	3	0.84	0.43	0.84	0.84	0.91
	4	0.70	0.71	0.70	0.70	0.84
	5	0.96	0.13	0.96	0.96	0.99
RMSprop	6	0.77	1.74	0.77	0.77	0.84
	7	0.88	0.26	0.88	0.88	0.96
	28	0.94	0.15	0.94	0.94	0.98
	29	0.98	0.05	0.98	0.98	0.99
	30	0.97	0.06	0.97	0.97	0.99
	1	0.58	2.29	0.58	0.58	0.67
	2	0.73	0.50	0.73	0.73	0.87
	3	0.92	0.25	0.92	0.92	0.97
	4	0.94	0.20	0.94	0.94	0.98
Adam	5	0.81	0.47	0.81	0.81	0.89
	6	0.61	1.25	0.61	0.61	0.81
	7	0.96	0.10	0.96	0.96	0.99
	28	0.97	0.09	0.97	0.9722	0.9944
	29	0.97	0.08	0.98	0.9815	0.9953
	30	0.94	0.15	0.94	0.9450	0.9883

maximum accuracy of 97% with a minimum loss of 0.08. The results for the average accuracy of the validation and loss are also depicted in **Table 2**. Average Validation loss and Accuracy using RMSprop optimizer is 0.33 and 90% while 9.22 and 85% using Adam Optimizer. The graph representation for the same is given in **Figure 11**.

4.4 Comparison of the Proposed Work With the Existing Models

Table 3 shows a comparison of the proposed work with existing models which have been used for the recognition of biomedical images. The table shows the values of parameters obtained for the recognition of biomedical images. It can be observed in the

TABLE 2 | Comparison of the average validation loss and accuracy obtained by two optimizers.

Parameter/Optimizer		RMSprop	Adam
Average	Validation Loss	0.33	9.22
Average	Validation Accuracy	0.90	0.85

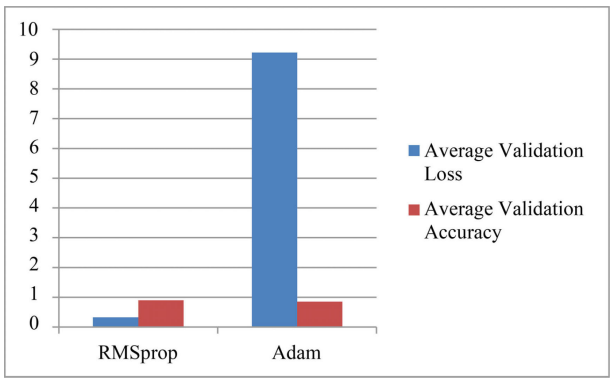


FIGURE 11 | Average Validation Values Obtained by using two Optimizer.

TABLE 3 | Comparison of the proposed work with existing models.

[Ref No.]/ Year	Worked on	Technique used	Parameters obtained
(6)/2020	Lungs Chest X-ray	• Manual feature extraction • Classifiers used: Random forest, logistic regression, Multilayer perceptron (MLP) technique	Accuracy: 95.39%
(7)/2020	CT Scan Images	• Manual feature extraction • Classifiers used: Random forest and logistic regression.	AUC:97%
(8)/2019	Chest X-ray	• Manual feature extraction • Classifiers used: Decision tree, Support vector machine (SVM), and logistic regression.	Accuracy: 94.5%
(13)/2021	Chest X-ray	• Manual feature extraction by employing multi-kernel local binary • Machine Learning based classifier	Accuracy: 97.01%.
(14)/2020	Chest X-ray	• Automatic Feature Extraction • Deep Learning based CNN Model	Accuracy: 90.68%.
(15)/2019	Chest X-ray	• Automatic Feature Extraction • Deep Learning based CNN Model	Accuracy: 93.73%.
(16)/2017	Chest X-ray	• Automatic Feature Extraction • DenseNet-121 CNN model	F1 Score: 76.8%
Proposed Model	Chest X-ray	• Automatic Feature Extraction • Deep Learning based CNN Model	Accuracy:98%

table that the highest recognition accuracy obtained by the proposed work is 98% by employing CNN based deep learning model in which automatic feature extraction takes place while others have achieved less accuracy.

5 CONCLUSION

As Pneumonia is a life-threatening disease that occurs due to an infection in the lungs if not detected at an early stage. Chest X-ray is the most common method for the diagnosis of the disease. Radiotherapist manually observes the features from the Chest X-ray which can help in the identification of the Pneumonia. A CNN model is proposed for the recognition of chest X-rays which are classified into 2 classes: Healthy X-rays and pneumonic X-rays. A CNN model employing two optimizers: RMSprop and Adam are used for recognition. Kaggle dataset is used for the recognition having 5000 samples of images for each class. From the obtained results, it has been observed that the RMS prop optimizer with an LR of 0.01 has helped in obtaining the maximum validation accuracy of 98% while the Adam optimizer has achieved the maximum validation accuracy of 97%. In terms of the average validation values obtained using two optimizers, it has been found that the highest average accuracy and the lowest average loss of 90% and 0.33 are also achieved using RMSprop Optimizer for the available dataset. Huge data are being taken into consideration for recognition and classification. The future direction could be to think about reducing the processing time of the data. Selective information processing is the solution, which means processing the required part of the image only instead of the whole image.

It can be concluded from the obtained results that RMSprop has outperformed Adam Optimizer for the recognition of the available dataset.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Conceptualization, SS, SG, DG. Data curation, SS, SG, DG, JR, SJ, JK, ME. Formal analysis, SS, SG, DG, JR, SJ. Funding acquisition, JK and JR. Investigation, SS, SG, DG, JR, SJ. Methodology, SS, SG, DG, JR, SJ, JK. Project administration, SG and DG. Writing—original draft, SS, SG, DG, JR, SJ. Writing—review and editing, SS, SG, DG, JR, SJ, JK, ME. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by the Technology Development Program of MSS [No. S3033853] and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A4A1031509).

ACKNOWLEDGMENTS

Our deepest gratitude goes to the anonymous reviewers for improving the quality of the paper.

REFERENCES

1. WHO. *Pneumonia*. World Health Organization (2019). Available at: https://www.who.int/news-room/fact-sheets/detail/pneumonia.
2. Neuman M, Lee E, Bixby S, Diperna S, Hellinger J, Markowitz R. Variability in the Interpretation of Chest Radiographs for the Diagnosis of Pneumonia in Children. *J Hosp Med* (2012) 7:294–8. doi: 10.1002/jhm.955PMID:22009855
3. Williams G, Macaskill P, Kerr M, Fitzgerald D, Isaacs D, Codarini M. Variability and Accuracy in Interpretation of Consolidation on Chest Radiography for Diagnosing Pneumonia in Children Under 5 Years of Age. *Pediatr Pulmonol* (2013) 48:1195–200. doi: 10.1002/ppul.22806PMID:23997040
4. Albahli S, Rauf HT, Algosaibi A, Balas VE. AI-Driven Deep CNN Approach for Multi-Label Pathology Classification Using Chest X-Rays. *PeerJ Comput Sci* (2021) 7:e495. doi: 10.7717/peerj-cs.495
5. Albahli S, Rauf H, Arif M, Nafis M, Algosaibi A. Identification of Thoracic Diseases by Exploiting Deep Neural Networks. *Neural Netw* (2021) 5:6. doi: 10.32604/cmc.2021.014134
6. Chandra T, Verma K. Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: *Proceedings of 3rd International Conference on Computer Vision and Image Processing* (India: Springer) (2020). p. 21–33.
7. Yue H, Yu Q, Liu C, Huang Y, Jiang Z, Shao C. Machine Learning-Based CT Radiomics Method for Predicting Hospital Stay in Patients With Pneumonia Associated With SARS-CoV-2 Infection: A Multicenter Study. *Ann Trans Med* (2020) 8:1–7. doi: 10.21037/atm-203026PMID:32793703
8. Kuo K, Talley P, Huang C, Cheng L. Predicting Hospital-Acquired Pneumonia Among Schizophrenic Patients: A Machine Learning Approach. *BMC Med Inf Decision Making* (2019) 19:1–8. doi: 10.1186/s12911-019-0792-1
9. Meraj T, Hassan A, Zahoor S, Rauf H, Lali M, Ali L. Lungs Nodule Detection Using Semantic Segmentation and Classification With Optimal Features. *Preprints* (2019) 10:4135–49. doi: 10.20944/preprints201909.0139.v1
10. Rajinikanth V, Kadry S, Damas ˇevičius R, Taniar D, Rauf H. Machine-Learning- Scheme to Detect Choroidal-Neovascularization in Retinal OCT Image. In: *2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)* (Chennai, India: IEEE. (2021). p. 1–5.
11. Kadry S, Nam Y, Rauf H, Rajinikanth V, Lawal I. Automated Detection of Brain Abnormality Using Deep-Learning-Scheme: A Study. In: *2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)* (Chennai, India: IEEE) (2021). p. 1–5.
12. Rajinikanth V, Kadry S, Taniar D, Damas ˇevičius R, Rauf H. Breast-Cancer Detection Using Thermal Images With Marine-Predators-Algorithm Selected Features. In: *2021 Seventh International Conference on Bio Signals, Images, And Instrumentation (ICBSII)* (Chennai, India: IEEE) (2021). p. 1–6.
13. Tuncer T, Ozyurt F, Dogan S, Subasi A. A Novel Covid-19 and Pneumonia Classification Method Based on F-Transform. *Chemometr Intell Lab Syst* (2021) 210:104256. doi: 10.1016/j.chemolab.2021.104256
14. Sharma H, Jain J, Bansal P, Gupta S. Feature Extraction and Classification of Chest X-Ray Images Using Cnn to Detect Pneumonia. In: *2020 10th International Conference On Cloud Computing, Data Science & Engineering (Confluence)* (Nodia, India: IEEE) (2020). p. 227–31.
15. Stephen O, Sain M, Maduh U, Jeong D. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *J Healthcare Eng* (2019) 2019:1–7 . doi: 10.1155/2019/4180949
16. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T. *Chexnet: Radiologist-Level Pneumonia Detection on Chest X-Rays With Deep Learning*. ArXiv Preprint (2017).
17. Juneja S, Dhiman G, Kautish S, Viriyasitavat W, Yadav K. A Perspective Roadmap for IoMT-Based Early Detection and Care of the Neural Disorder, Dementia. *J Healthcare Eng* (2021) 2021 1–11. doi: 10.1155/2021/6712424. Hindawi Limited.
18. Sharma S, Gupta S, Gupta D, Juneja S, Singal G, Dhiman G, et al. Recognition of Gurmukhi Handwritten City Names Using Deep Learning and Cloud Computing. *Sci Program* (2022) 2022:1–16. doi: 10.1155/2022/5945117
19. Dhankhar A, Juneja S, Juneja A, Bali V. Kernel Parameter Tuning to Tweak the Performance of Classifiers for Identification of Heart Diseases. *Int J E-Health Med Commun* (2021) 12(4):1–16. doi: 10.4018/IJEHMC.20210701.oa1
20. Kumar N, Gupta M, Gupta D, Tiwari S. Novel Deep Transfer Learning Model for COVID-19 Patient Detection Using X-Ray Chest Images. *J Ambient Intell Humaniz Comput* (2021) 2021:1–10. doi: 10.1007/s12652-021-03306-6
21. Dhiman G, Juneja S, Viriyasitavat W, Mohafez H, Hadizadeh M, Islam MA, et al. A Novel Machine-Learning-Based Hybrid CNN Model for Tumor Identification in Medical Image Processing. *Sustainability (Switzerland)* (2022) 14(3):1–13. doi: 10.3390/su14031447
22. Dhiman G, Rashid J, Kim J, Juneja S, Viriyasitavat W, Gulati K. Privacy for Healthcare Data Using the Byzantine Consensus Method. *IETE J Res* (2022) 2022:1–12. doi: 10.1080/03772063.2022.2038288
23. Chowdhary CL, Alazab M, Chaudhary A, Hakak S, Gadekallu TR. *Computer Vision and Recognition Systems Using Machine and Deep Learning Approaches: Fundamentals, Technologies and Application*. The Institution of Engineering and Technology (2021).
24. Ali TM, Nawaz A, Rehman AU, Ahmad RZ, Javed AR, Gadekallu TR, et al. A Sequential Machine Learning Cum Attention Mechanism for Effective Segmentation of Brain Tumor. *Front Oncol* (2022) 2129. doi: 10.3389/fonc.2022.873268
25. Mannan A, Abbasi A, Javed AR, Ahsan A, Gadekallu TR, Xin Q. Hypertuned Deep Convolutional Neural Network for Sign Language Recognition. *Comput Intell Neurosci* (2022) 2022:1–10. doi: 10.1155/2022/1450822
26. Sharma S, Gupta S, Gupta D, Juneja S, Gupta P, Dhiman G, et al. Deep Learning Model for the Automatic Classification of White Blood Cells. *Comput Intell Neurosci* (2022) 2022:1–13 . doi: 10.1155/2022/7384131
27. Rahman T, Chowdhury M, Khandakar A, Islam K, Islam K, Mahbub Z. Transfer Learning With Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-Ray. *Appl Sci* (2020) 10:3233. doi: 10.3390/app10093233

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sharma, Gupta, Gupta, Rashid, Juneja, Kim and Elarabawy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fpubh.2022.925147

A Study on the Correlation Between Age-Related Macular Degeneration and Alzheimer's Disease Based on the Application of Artificial Neural Network

Meng Zhang[1], *Xuewu Gong*[2], *Wenhui Ma*[3], *Libo Wen*[4], *Yuejing Wang*[1] *and Hongbo Yao*[1]*

[1] Histology and Embryology Section, Qiqihar Medical University, Qiqihar, China, [2] The Second Affiliated Hospital of Qiqihar Medical University, Qiqihar, China, [3] Computer Experimental Teaching Center, Qiqihar Medical University, Qiqihar, China, [4] Physiology Section, Qiqihar Medical University, Qiqihar, China

OPEN ACCESS

Edited by:
Thippa Reddy Gadekallu,
VIT University, India

Reviewed by:
Vinit Gunjan,
CMR Institute of Technology, India
D. S. Bhupal Naik,
Vignan's Foundation for Science, Technology and Research, India

***Correspondence:**
Hongbo Yao
22868204@qq.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 21 April 2022
Accepted: 11 May 2022
Published: 30 June 2022

Citation:
Zhang M, Gong X, Ma W, Wen L, Wang Y and Yao H (2022) A Study on the Correlation Between Age-Related Macular Degeneration and Alzheimer's Disease Based on the Application of Artificial Neural Network. Front. Public Health 10:925147. doi: 10.3389/fpubh.2022.925147

Age-related Macular Degeneration (AMD) is a kind of irreversible vision loss or disease caused by retinal pigment epithelial cells and neuroretinal degeneration, which has become the main cause of vision loss and blindness of the elderly over 65 years old in developed countries. The main clinical manifestations are cognitive decline, mental symptoms and behavioral disorders, and the gradual decline of daily living ability. In this paper, a feature extraction method of electroencephalogram (EEG) signal based on multi-spectral image fusion of multi-brain regions is proposed based on artificial neural network (ANN). In this method, the brain is divided into several different brain regions, and the EEG signals of different brain regions are transformed into several multispectral images by combining with the multispectral image transformation method. Using Alzheimer's disease (AD) classification algorithm, the depth residual network model pre-trained in ImageNet was transferred to sMRI data set for fine adjustment, instead of training a brand-new model from scratch. The results show that the proposed method solves the problem of few available medical image samples and shortens the training time of ANN model.

Keywords: artificial neural network, age-related macular degeneration, Alzheimer's disease, relevance, correlation

INTRODUCTION

Alzheimer's disease (AD) is a kind of neurodegenerative disease, which is mainly characterized by memory disorder, cognitive disorder and language disorder, and is the most common cause of senile dementia (1). The latest survey report released by the International Association of AD shows that a new AD case will appear every 3 s around the world. The exact etiology and pathogenesis of AD are not clear, so there is no complete cure for AD. Once suffering from AD, it will cause destructive damage to the physiological structure of the brain. For example, various brain regions will gradually shrink. Because this injury is irreversible, the life expectancy of patients diagnosed with AD is only about 7 years (2). Therefore, how to control the progress of AD is extremely important.

Age-related macular degeneration (AMD) is the main cause of blindness in the elderly, especially irreversible visual loss. Cognitive function is an important part of the higher cerebral cortex, which is composed of memory, attention, calculation, orientation, executive ability, and so on. AD is the most common disease type of cognitive dysfunction. The prevalence of AMD and cognitive dysfunction increases with age, which has a great impact on people's life, study and work (3, 4). Literature (5) found that the score of the simple mental state examination scale in AMD group was lower than that of the matched control group without AMD, such as age, sex, and years of education, and the prevalence of AD in AMD patients was higher than that in the control group (40.7 vs. 20.4%, $p = 0.03$). Literature (6) a large-scale prospective study of AD, AMD, and control group showed that there was no significant difference between AD group and control group. Some studies have found that AMD and AD are not only common in epidemiology, but also common in molecular aspects (7–10). It is speculated that oxidative stress may be the earliest feature of AD, and the level of heme oxygenase-1, a sensitive marker of oxidative stress, was found in the brain tissues of AD and patients with mild cognitive impairment. Literature (11) found that aggregated Aβ is not only the activator of complement system in brain, but also the activator of microglia, which leads to the increase of oxidative stress. It has been found that microglia activation also exists around drusen and subretinal space, causing damage to brain and eyes. Literature (12, 13) in a population-based prospective cohort study, it was found that the newly diagnosed dry and wet AMD group was more likely to suffer from AD or Alzheimer's disease ($p = 0.044$) than the control group with no AMD at all and matched age, sex, and entry time. Compared with wet AMD patients, the relationship between dry AMD patients and AD is more obvious.

In the deep learning algorithm, convolutional neural network and deep neural network are used to classify electroencephalogram (EEG) signals of AMD and NC. Literature (14) classifies resting EEG signals of AD patients, patients with mild cognitive impairment and normal people by combining convolutional neural network with multilayer feed-forward perceptron, and finally obtains the best accuracy of 85% for the second classification and 82% for the third classification. Literature (15) classified the resting EEG signals of patients with AD, patients with mild cognitive impairment, and normal people by using convolutional neural network, and finally got the best accuracy of 89.8% for the second classification and 83.3% for the third classification (16). In the visual working memory paradigm, features are extracted from patients' task EEG signals, and then multi-kernel SVM is used for classification. Finally, the maximum classification accuracy is 91.76 and 82.23%, respectively, under the stimulus of 0-back and 1-back tasks.

Artificial neural network (ANN) is a complex network system, which consists of a large number of simple basic elements-neurons connected with each other, and performs information parallel processing and non-linear transformation by simulating the way of human brain nerve processing information (17). ANN can simulate many basic functions and ways of thinking of the brain, and acquire external knowledge through learning and store it in the network. ANN has been widely used in economy, engineering, medicine, biology, and other fields, and has become a unique information processing discipline. Compared with digital computers, ANN is closer to the human brain in terms of composition principle and functional characteristics. "Instead of executing operations step by step according to a given program, ANN can adapt itself to the environment and summarize laws!" Complete some operation, identification or process control (18). Based on the basic theory of ANN, this paper proposes an AD classification algorithm based on transfer learning and deep residual network. Migrate the pre-trained depth residual network model in ImageNet to sMRI data set for fine-tuning, instead of training a brand-new model from scratch; The results show that the proposed method solves the problem of few available medical image samples.

CORRELATION MECHANISM BETWEEN AMD AND AD

Although AMD and AD are degenerative diseases of different tissues, because retina is an integral part of central nervous system, there may be a correlation between the pathology and pathogenesis of these two diseases.

Pathophysiological Research

Some studies have found that AMD and AD are not only common in epidemiology, but also common in molecular aspects (19). For example, the characteristic senile plaque of AD and the marker drusen of AMD have common active ingredients, the most important of which is Aβ, both of which contain and (20).

In addition, a study showed that the defects of electroretinogram were eliminated in a dose-dependent manner by administering antibodies targeting and terminals to mice. The decrease of Aβ level in retinal pigment epithelial deposits and the preservation of retinal pigment epithelial structure were related to anti-q antibody immunotherapy and visual protection. These observations are consistent with the decrease of amyloid and the improvement of cognitive function in AD mice treated with anti-Aβ antibody (21).

Relevance Mechanism

Chronic Inflammatory Reaction

Inflammatory reaction is the rapid response of cells to danger, with the aim of initiating immune response. In the short term, inflammatory reaction is beneficial, but a long-term chronic inflammatory reaction is harmful to the body. Long-term inflammation is related to the development of various chronic diseases, such as autoimmune diseases and neurodegenerative diseases (22). Amyloid protein and lipofuscin increase with age in normal eyes and brain, which activate inflammasome, complement system and autophagy lysosome and accelerate the progress of diseases (23). The studies have found that microglia activation also exists around drusen and subretinal space (24), which causes damage to brain and eyes. These phenomena are both the causes and consequences of Aβ formation. In this way, an irreversible positive feedback mechanism of pro-inflammatory cytokines and protease secretion occurs, which drives the progress of the disease (25).

Oxidative Stress

In the brain and eyes, the function of physiological cells needs oxidative stress, but when it exceeds a certain threshold, it will be cytotoxic. However, the human body has an antioxidant system to rebuild the internal balance: the dysfunction of this system may contribute to AD pathophysiology. It is speculated that oxidative stress may be the earliest feature of AD, and the level of heme oxygenase-1, a sensitive marker of oxidative stress, was found in the brain tissues of AD and patients with mild cognitive impairment. It has also been found that autophagy and lysosomes also participate in the oxidative stress process in the eyes of patients with AMD and the brains of patients with AD (26). In a word, excessive oxidative stress and dysfunction of mitochondria and lysosomes seem to be common pathophysiology in the pathogenesis of AMD and AD.

RESEARCH METHOD

Artificial Neural Network

Neural network is composed of a large number of neurons connected with each other. The processing ability of a single neuron is not strong, but after a large number of neurons are interconnected into a network, its processing ability is greatly improved. Neural network is a model obtained by abstracting and simplifying biological neural network on the basis of simulating the basic characteristics of brain. In neural networks, the storage of knowledge and information is characterized by the interconnected and distributed physical connections among neurons, while learning and recognition are the dynamic changes of the weighting coefficients of connections among neurons.

It is the basic information processing unit of ANN operation, and the neuron model contains 3 basic elements: synapse, adder and activation function. **Figure 1** shows the neuron model, where the input is signal and represents the synaptic weight of neuron .

The neuron in **Figure 1** can be described by the following equations:

$$u_j = \sum_{i=1}^{n} \left(w_{nj} x_n\right) \tag{1}$$

$$y_j = \phi\left(u_j + b_j\right) \tag{2}$$

At present, according to different problems, different ANN models are proposed, and each algorithm is different. Common network models include BP neural network model, Hofield neural network model, linear neural network model, and Kohonen self-organizing model. These ANNs are not suitable for all the problems, so we must choose the appropriate ANNs according to the different problems to be solved. BP neural network and radial basis function neural network are often used in medical diagnosis and prediction, and perceptron model is mainly used for classification in clinic.

Multi-Brain Multispectral Image Fusion Algorithm

A large number of studies have shown that there are differences in EEG signals between AMD and Normal Control (NC), which are generally manifested in five brain regions. The five brain regions are frontal lobe, parietal lobe, left temporal lobe, right temporal lobe, and occipital lobe (27). For example, in frontal lobe, there are differences between theta frequency band and delta frequency band between AMD and NC group. In parietal lobe, AMD and NC have different phase coupling in the cross-frequency band between alpha frequency band and beta frequency band. In this study, EEG signals of five brain regions were converted into multispectral images by multispectral image conversion method. Multispectral images representing different brain regions are obtained by using the method of multispectral image transformation alone in five brain regions.

In this study, the EEG signals of five brain regions were transformed into five multispectral images by the method of multispectral image conversion. Each brain region contains more electrodes, and if all the electrodes are used, the amount of calculation will be greatly increased. Therefore, some electrodes are selected from each brain region, and the EEG signals of these electrodes are converted into multispectral images.

In this study, pixel-level weighted average image fusion algorithm is used to fuse multispectral images of multiple brain regions. The fusion formula of the weighted average fusion method is shown in formula (3):

$$Y = \sum_{n} A_n X_n \tag{3}$$

In formula (3), A_n represents the corresponding weight, $\sum_n A_n = 1$, X_n represents the corresponding pixel, and Y represents the generated new pixel.

Because the multispectral image formed in this study is a 40*40*3 color image, each pixel X is composed of 3 values (RGB 3 primary colors), namely $X = (X_R, X_G, X_B)$. According to the above weighted average fusion formula, the following formula is obtained.

$$\begin{aligned} Y_R &= \sum_{n} A_n X_R \\ Y_G &= \sum_{n} A_n X_G \\ Y_B &= \sum_{n} A_n X_B \end{aligned} \tag{4}$$

That is, for the pixels in each image, the corresponding R, G, and B are taken for weighted average fusion, and finally the corresponding pixels of the new multispectral image are generated.

$$Y = \left(\sum_{n} A_n X_R, \sum_{n} A_n X_G, \sum_{n} A_n X_B\right) \tag{5}$$

In this study, the pixel-level weighted fusion method of image fusion algorithm is used to fuse these 5 images to form a new image, as shown in **Figure 2**.

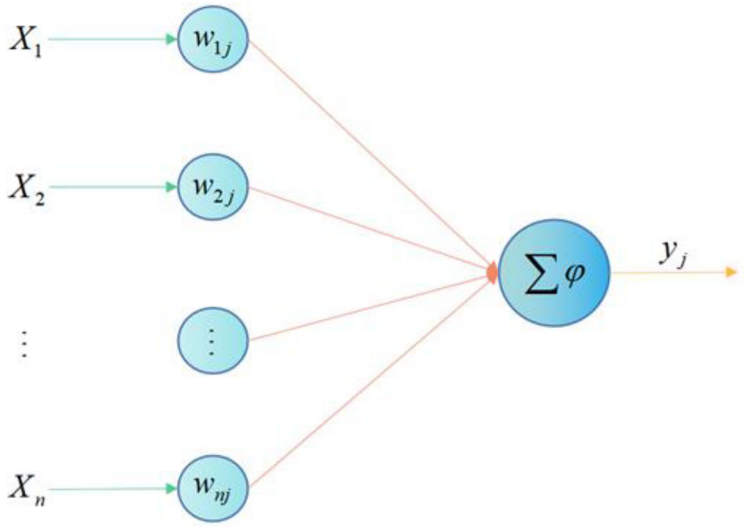


FIGURE 1 | Neuron model.

The feature extraction method of EEG signal based on multispectral image fusion of brain regions is a combination of multispectral image transformation method and image fusion method. Firstly, the EEG signals of each brain region are extracted according to the electrode position (28). Secondly, the multi-spectral image transformation method is used to transform the EEG signals of each brain region into multi-spectral images of the characteristics of each brain region. Finally, the image weighted average fusion algorithm is used to fuse the multispectral images of the five brain regions to form the fused multispectral image as shown in **Figure 2**.

Multi-brain multispectral image fusion;

Input: EEG of task EEG.

Output: multispectral images of multiple brain regions and fused multispectral images.

BEGIN

(1) $X_i = EEG(i)$. //The electrode signals of the 5 selected brain regions are extracted, in which the values of i are e, d, ZN, YN, and z, which respectively represent frontal lobe, parietal lobe, left temporal lobe, right temporal lobe and occipital lobe.

(2) $Y_i = FFT(X_i)$. //Fourier transform was carried out on EEG signals of each brain region.

(3) $Z_j = Y_i(j)$. //The values of j are alpha, beta, and theta, which means extracting signals of 3 frequency bands.

(4) $Z_j = square(Z_j)$. //Take the sum of squares for the signals of the three frequency bands respectively.

(5) Local_2d = AEP(Local_3d). //According to AEP method, the electrode 3d coordinates are converted into 2d coordinates.

(6) $D_i = Conver_to_image(Z_{theta}, Z_{theta}, Z_{beta} = R, G, B \mid$ Local _2d, Clough-Tocher$)$. //The features of the sum of squares of the three frequency bands of each electrode are transformed into images according to the 2d coordinates of the electrodes, and the surrounding pixels are filled up by Clough-Tocher interpolation method. The images formed by the three frequency bands are used as RGB channels of color images, and a color multispectral image is formed.

(7) $Output(D_i)$. //Outputting multispectral images of five brain regions.

(8) $D = \sum_i W_i D_i$. //Image-level weighted average fusion method, that is, adding corresponding pixels, in which the weight sum is 1, that is, $\sum_i W_i = 1$.

(9) $Output(D)$. //Output a new multispectral image after fusion.

END

Group Diagram Construction

The key to construct population map is to select the phenotypic measure that best explains the similarity between imaging data or the similarity between subject tags. In the ANN method, there are h sets of phenotypic information $M = \{M_h\}$ (such as gender and age of the subjects), and the adjacency matrix W of the population graph is defined as:

$$W(v, w) = sim(A_v, A_w) \sum_{h=1}^{H} \gamma(M_h(v), M_h(w)) \quad (6)$$

This formula represents the edge weight between the tested v and the tested w, which consists of 2 parts: $sim(A_v, A_w)$ represents the similarity of brain network features between the tested, and increases the edge weight between the most similar nodes; γ represents the similarity between phenotypic information.

Specifically, feature similarity is defined as:

$$sim(A_v, A_w) = \exp\left(-\frac{(\rho(x(v), x(v)))^2}{2\sigma^2}\right) \quad (7)$$

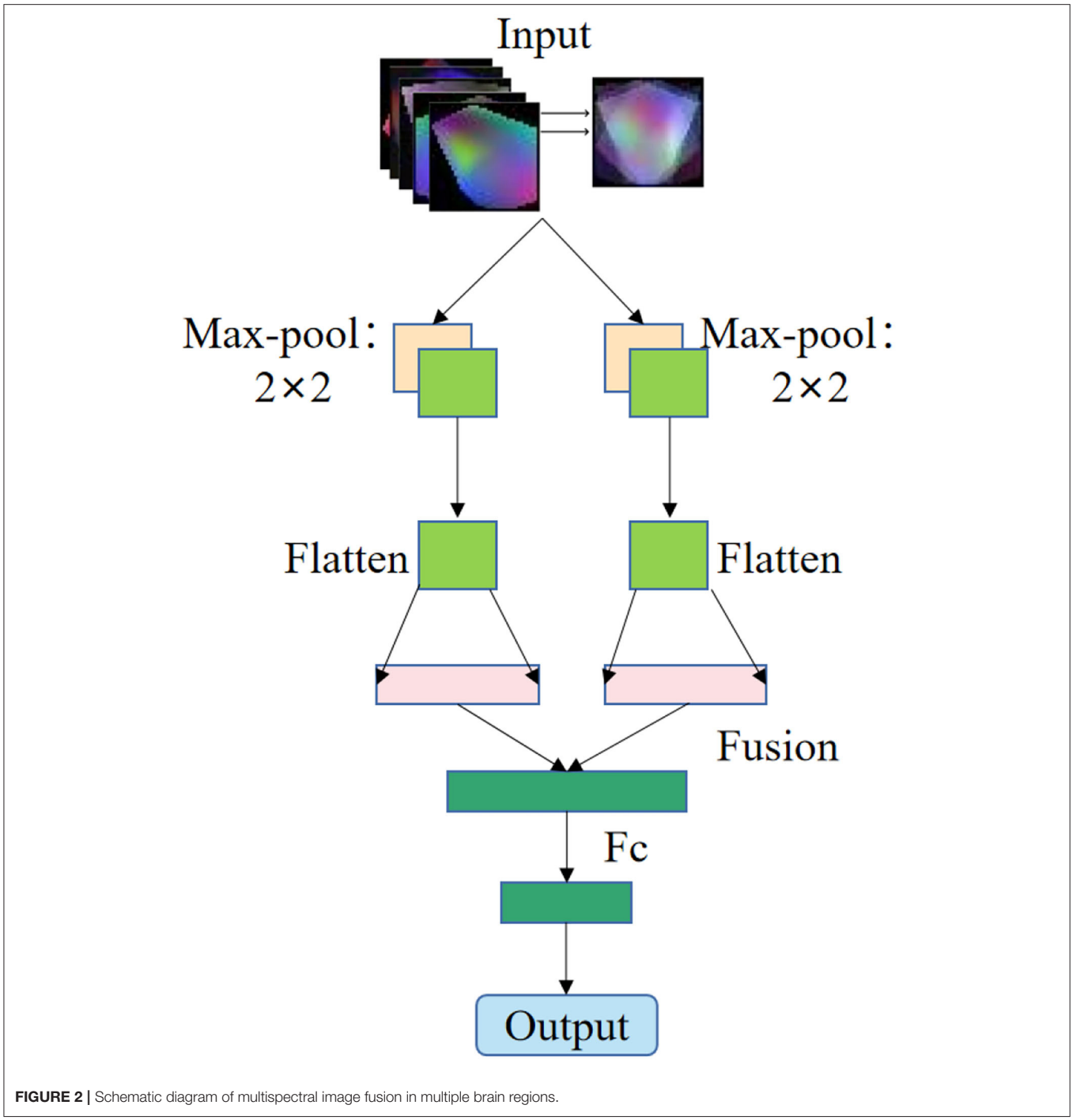


FIGURE 2 | Schematic diagram of multispectral image fusion in multiple brain regions.

Where, $\rho\left(x\left(v\right),x\left(v\right)\right) \in (0,1)$ is the similarity distance between two subjects, and the more similar the two groups are, the closer the similarity distance is to 1; σ is the kernel width, so the neighborhood range can be customized.

The idea of this feature similarity measure is that subjects belonging to the same class (healthy or ill) have more similar networks than subjects from different classes, that is, larger sim values. The definition of phenotypic information similarity can be divided into quantitative type and qualitative type (29). For qualitative phenotypic information with gender as an example, the Kroneckerdelta formula is used to defineγ. For example, if two subjects have the same gender, the weight of the edge between two nodes will increase.

For quantitative phenotypic information with age as an example, if the difference between them is less than the present value, it is considered to be similar, and the edge weight between

two nodes is set to 1; otherwise, it is considered that the information is not similar, and the edge weight is 0. γ is defined by a unit step function with threshold θ, which includes:

$$\gamma\left(M_h\left(v\right), M_h\left(w\right)\right) = \begin{cases} 1, & \left(\left|M_h\left(v\right), M_h\left(w\right)\right| < \theta\right) \\ 0, & \text{Otherwise} \end{cases} \tag{8}$$

Finally, the similarity of eigenvalues is multiplied by the similarity of phenotypic information, which is integrated into the total edge weight to obtain the final group adjacency matrix, that is, the graph structure. This definition of similarity reflects strong similarity in the same test.

AD Classification Algorithm

The Residual Neural Network (ResNet) was proposed by He et al. (30). The depth ResNet consists of a group of residual blocks, each of which consists of convolution layer, batch normalization layer and active function rectification linear unit. ResNets can be regarded as multiple basic blocks connected in series with each other. This shortcut can skip each basic block in parallel and then connect it to the output of the stack layer. It is easier to optimize the residual mapping than to optimize the original, unreferenced mapping.

Based on the special design of deep ResNet, the problems such as gradient explosion and gradient disappearance are effectively suppressed, so that the network can be fully trained, the number of network layers is deepened, and the performance of parameter optimization space is improved. Because the pre-trained ResNet-20 model accepts an image input size of $214 \times 214 \times 4$, we convert each sMRI slice image into a size of $214 \times 214 \times 4$, and then subtract the average intensity value from the corresponding single image channel to normalize the image. Then, we will migrate the ResNet-20 model pre-trained in ImageNet to the data set for fine-tuning.

In addition, because the number of parameters in the fully connected layer is too large to be trained, we removed the original fully connected layer and retrained a fully connected layer $F\left(e\right)$ whose length depends on the number of categories, and e represents the number of categories.

In order to prevent over-fitting, we also added a dropout layer with a loss rate of 0.1. p_{ave} indicates that average pooling is used in this paper. The softmax function used in the training process is shown in the following formula:

$$\text{softmax}\left(\mathrm{F}_j\right) = \frac{\exp^{\mathrm{F}_j}}{\sum_{k=1}^{e} \exp^{\mathrm{F}_k}}, j = 1, 2, \cdots, e \tag{9}$$

We use the cross-entropy loss function to measure the performance of the classification model, and its output is a probability value between 0 and 1. Cross entropy loss increases with the distance that the predicted output deviates from the actual label. If the number of classes is two, the binary cross entropy loss is calculated as follows:

$$L\left(y, p\right) = -\left(y \log p + \left(1 - y\right) \log\left(1 - p\right)\right) \tag{10}$$

If $e > 2$, the classification cross entropy loss is calculated as follows:

$$L\left(y, p\right) = -\sum_{\mathrm{c}=1}^{\mathrm{e}} \mathrm{y}_{\mathrm{o,c}} \log\left(p_{\mathrm{o,c}}\right) \tag{11}$$

Where y is the actual value and p is the predicted value. The optimization algorithm for updating parameters is RMSprop (root mean square propagation), and the learning rate is set to 0.0001. The flow chart of AD classification algorithm based on transfer learning and deep residual network is shown in **Figure 3**.

RESULT ANALYSIS AND DISCUSSION

In this study, k-fold cross-validation method ($K = 5$) is used to evaluate the model, but the problem of this study belongs to classification. For the algorithm proposed in this study, the generalization performance of the model will be evaluated by selecting the functions of calculation accuracy, recall rate, F1 value, AUC value, average validation accuracy and average validation loss. Among them, for the new algorithm proposed in this study and its comparative experiments, all evaluation indexes are used, while for some basic experiments, only verification accuracy is used as the only evaluation index.

Experimental Results of EEG Feature Extraction Method Based on Multispectral Image Fusion in Multiple Brain Regions

Figure 4 shows the graph of the average verification accuracy of multispectral image experiments in multiple brain regions. YN stands for right temporal lobe, ZN stands for left temporal lobe, Z stands for occipital lobe, D stands for parietal lobe, 18 stands for 18 brain regions commonly used in related research, E stands for frontal lobe, and All stands for all brain regions selected by the above five regions.

It can be seen from the figure that the curve representing All is the highest, and the accuracy reaches the highest point around the 30th epochs, which can remain stable, but the curve fluctuates. In other cases, the curves representing E, 18 and D are slightly lower than All, and the accuracy reaches the highest point around the 35th epochs and remains stable. While the curves representing z, ZN and YN are low, and the accuracy reaches the highest point around the 50th epochs and remains stable.

Figure 5 shows the graph of the average verification loss function of multispectral image experiment in multiple brain regions.

It can be seen from **Figure 5** that the curve representing All is the lowest, and the loss function value can reach the lowest around the 30th epochs, but the curve fluctuates greatly. In other cases, the curves representing f, 18 and p are slightly higher than All, and the loss function value reaches the lowest point around the 50th epochs and remains stable. While the curves representing o, LT and RT are higher, and the loss function value reaches the lowest point around the 55th epochs and remains stable.

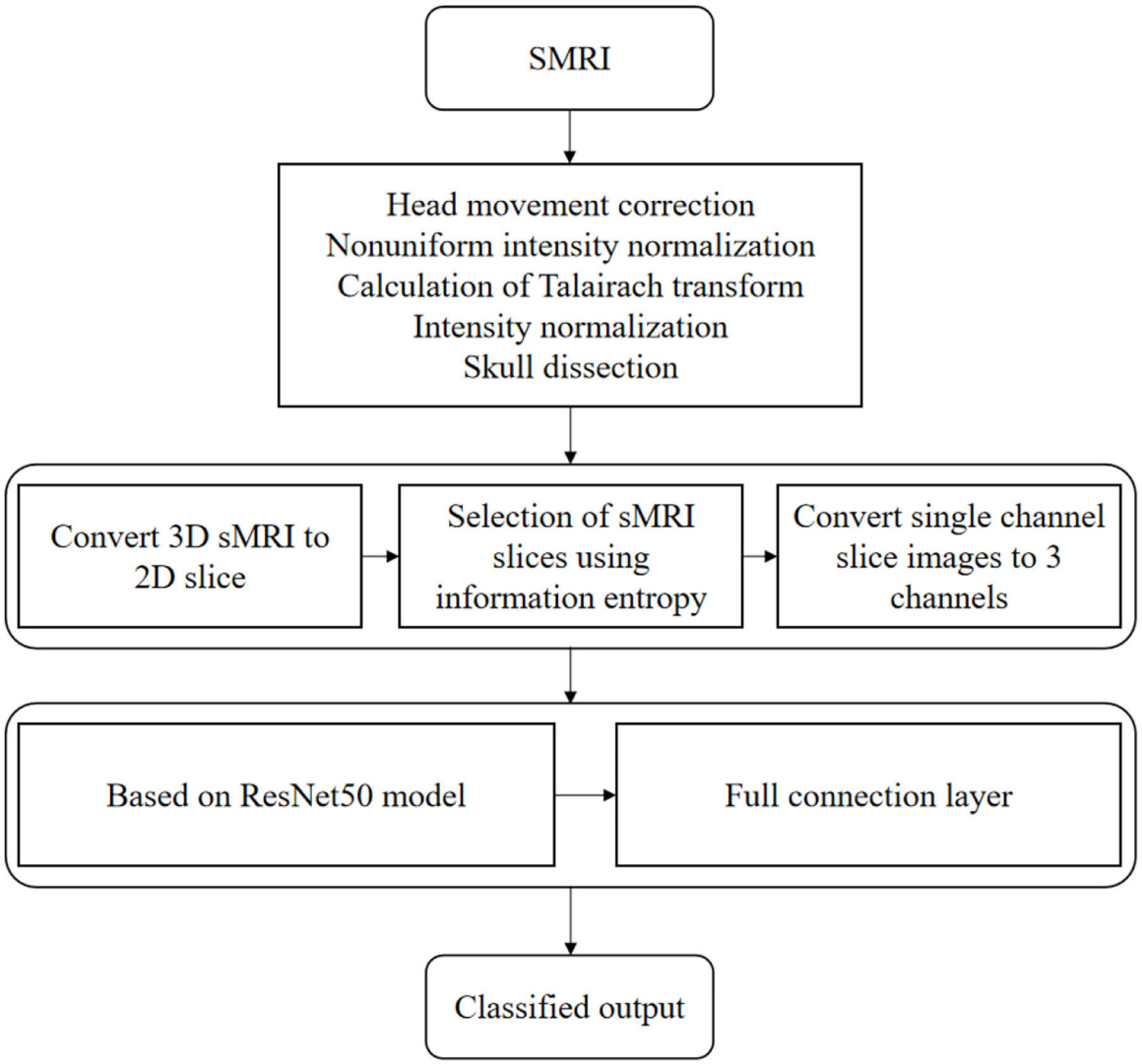


FIGURE 3 | AD classification algorithm flow chart.

Figure 6 shows the average verification accuracy rate, average recall rate, average F1 value and average AUC value of this experiment.

Figure 6 shows that the frontal lobe has the highest value among the 5 multi-spectral image classification performance indexes, which are 97.24% of the average verification accuracy, 97.80% of the average recall rate, 97.60% of the average F1 value and 99.01% of the average AUC value.

In addition, among the seven brain regions, the multispectral image classification performance index of ALL brain regions is in the best position, and the average verification accuracy rate, recall rate, F1 value and AUC value are 97.96, 97.82, 97.88, and 98.35%, respectively.

It can be concluded that the fused multispectral image can complement the information contained in the multispectral image of multi-brain region, so that the fused image has richer feature information, thus improving the classification accuracy. At the same time, the research results of this paper have surpassed the classification performance of previous AMD EEG studies.

In the future, many different image fusion technologies can be used to fuse multispectral images, and the positive and negative relationships among brain regions can also be considered, so as to refine the fusion of multispectral images with multiple brain regions, and the amount of data can be increased by increasing the number of subjects, so as to further verify the rationality of this method.

AD Classification Algorithm Analysis

All the classification models proposed in this paper are trained and tested in NVIDIA GeForce GTX 1060 GPU system based on Keras. One advantage of Keras is that it can easily and quickly prototype and run seamlessly on GPU. In this paper, transfer learning is introduced, only some layers of the pre-trained ANN model are trained, and the amount of data used in

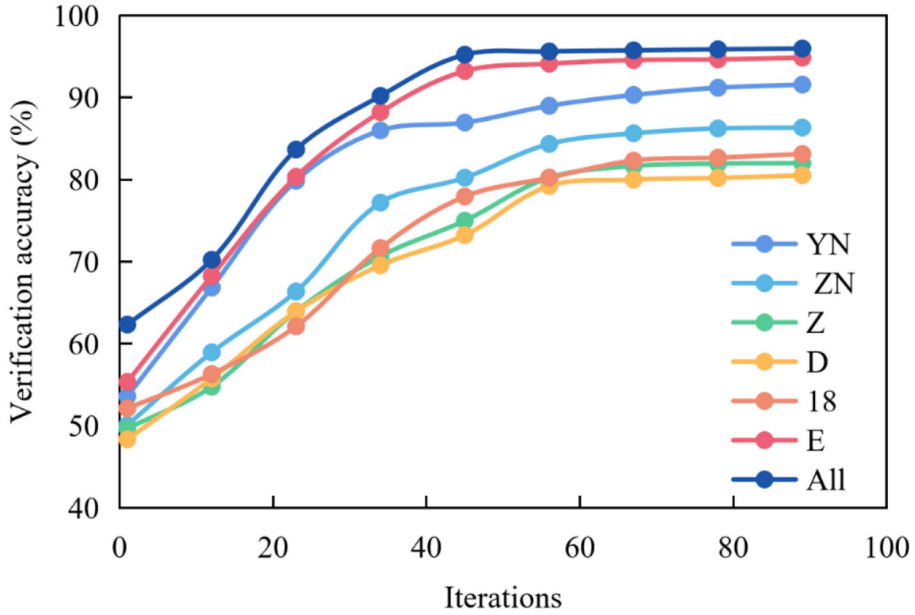


FIGURE 4 | Average verification accuracy curve of multispectral image fusion in multiple brain regions.

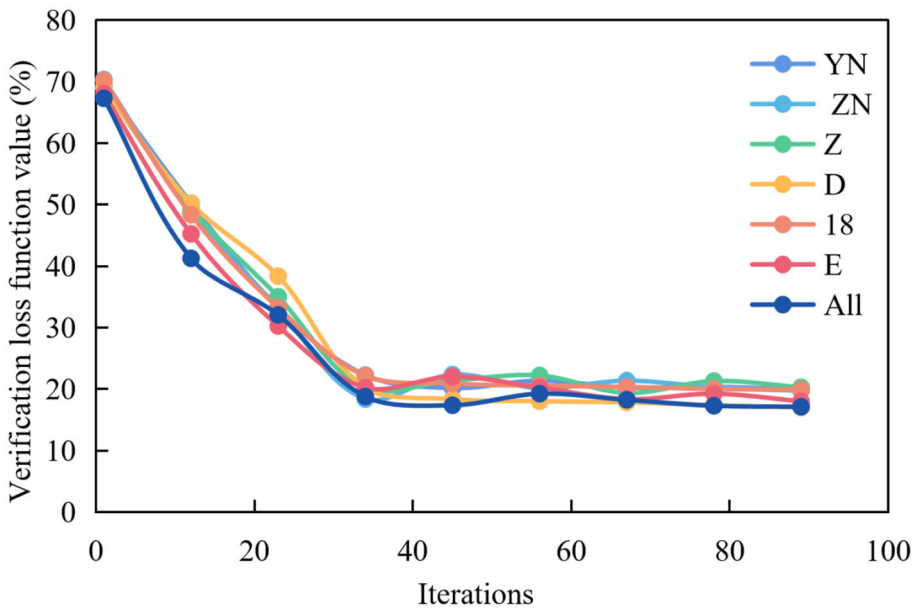


FIGURE 5 | Average verification loss function curve of multispectral image fusion in multiple brain regions.

this experiment is small, so this experiment can be completed in this GPU system.

In order to record the experimental results conveniently, we call our method ResNet-20_TL. The classification results of AD, NC, and AMD by the method proposed in this paper are shown in **Figure 7**.

For binary classification, the accuracy, sensitivity and specificity of the model trained in this paper are 97.66, 90.25, and 96.33% in AD/AMD classification, 90.01, 86.35, and 91.24% in AD/AMD classification, and 92.35 and 82.35% in NC/AMD classification, respectively. For ternary classification, the classification accuracy, sensitivity and specificity of the proposed method are 87.69, 73.21, and 85.46%, respectively.

It can be seen from **Figure 7** that AD and NC are the easiest categories to judge, while AMD is easy to be confused with AD

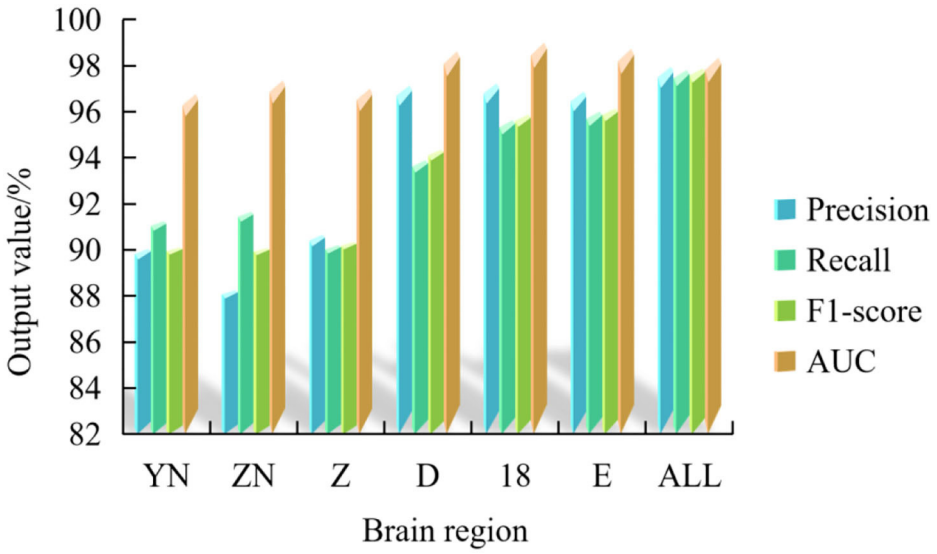


FIGURE 6 | Average accuracy rate, recall rate, F1 value and AUC of multispectral image fusion experiments in multiple brain regions.

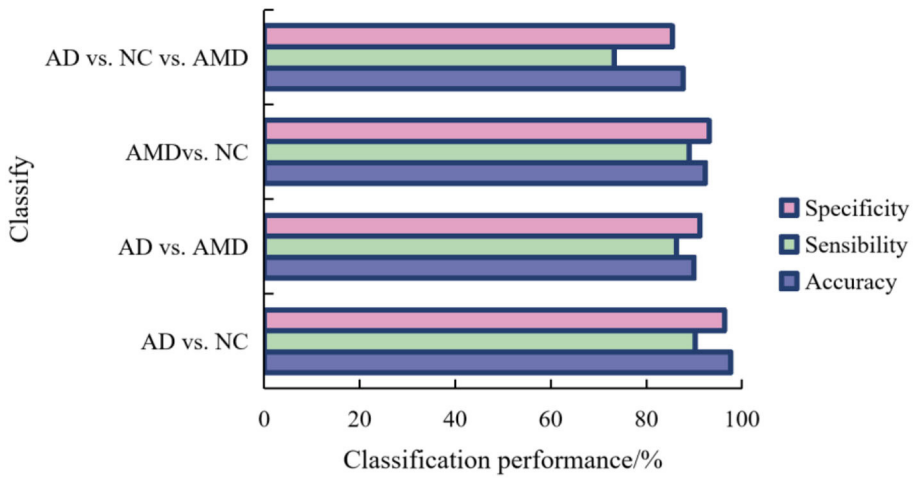


FIGURE 7 | PerformaNCe of ResNet-20_TL model in AD, NC, and AMD classification.

and NC. this is because the brain structure of the subjects in different stages of the disease changes to different degrees, the brain tissue of the subjects in AD shrinks seriously, and the brain of the subjects in NC is in a normal state, so the difference in scanned sMRI images is larger, and the proposed method is easier to extract discriminatory features, thus improving the classification performance.

It can also be seen from **Figure 7** that compared with binary classification, the accuracy and specificity of AD/NC/AMD classification are lower than 90%, and the sensitivity is lower, mainly because AMD is the intermediate stage between AD and AMD, which leads to the "distance" between features being not obvious enough, so it is not easy to distinguish, which leads to the unsatisfactory accuracy of ternary classification.

In addition, in order to prove the ability of migration learning, we also trained ResNet-20 network from scratch, which is represented by ResNet-20_scratch. For ResNet-20 model trained from scratch, 20 epochs were trained, and the batch was still 64. The adaptive learning rate was used, and the parameters were optimized by RMSProp.

In order to compare the classification performance of ResNet-20_TL model and ResNet-20_scratch model more intuitively, we use a column chart to show it, as shown in **Figure 8**. Obviously, the fine-tuned ResNet-20 model has better performance in each classification task than the ResNet-20 model trained from scratch.

In addition, facing the classification tasks of AMD/NC, AD/NC, AMD/AD, and AD/AMD/NC, the AUC comparison

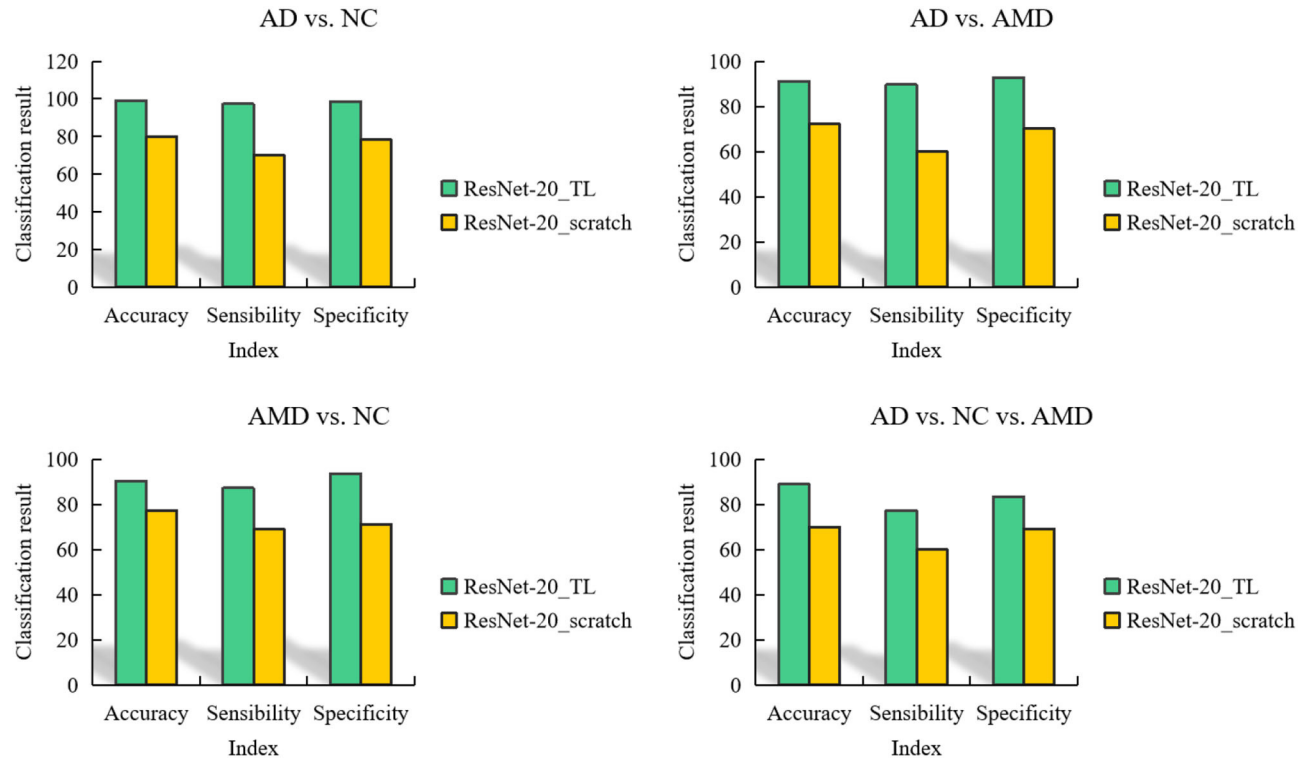


FIGURE 8 | Comparison of classification results obtained by 2 ANN models in 4 classification tasks.

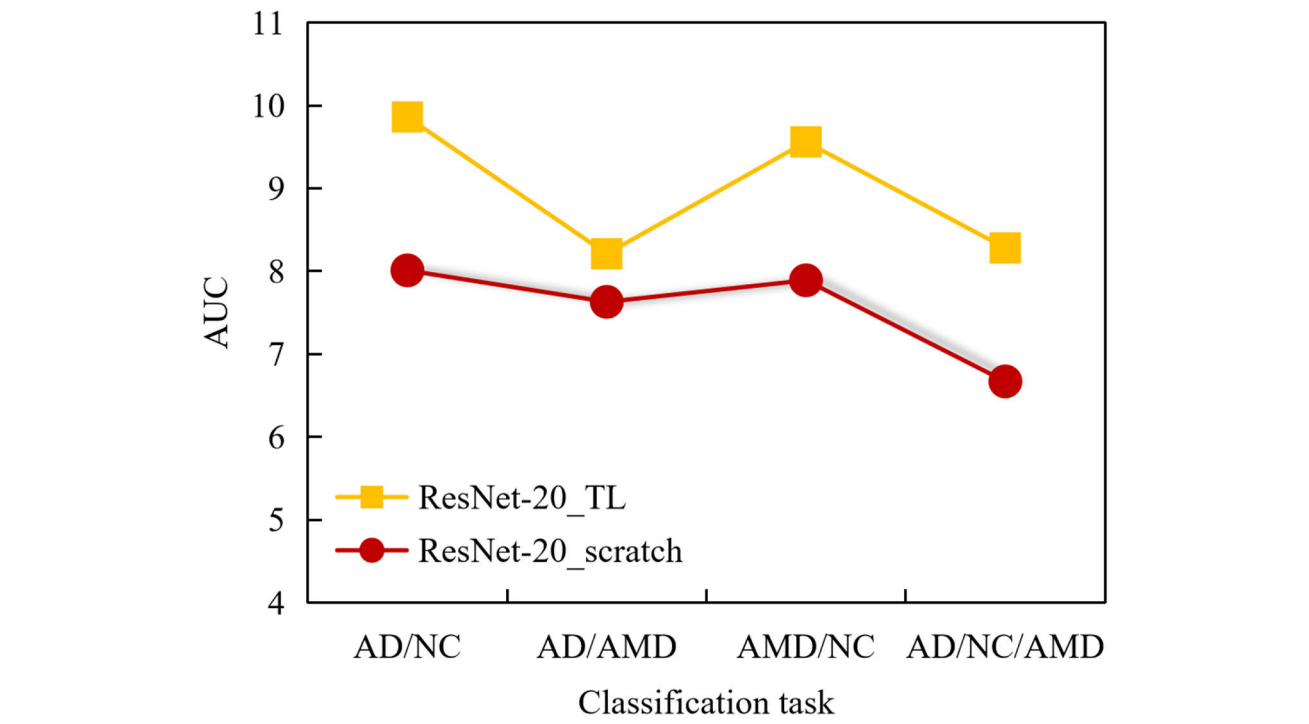


FIGURE 9 | AUC comparison curves of 2 ANN models in classification tasks AMD/NC, AD/NC, AMD/AD, and AD/AMD/NC.

diagram of the area under the curve obtained by the 2 ANN models is shown in **Figure 9**.

It can be seen from **Figure 9** that the AUC value of ResNet-20_scratch model trained from scratch is higher than 0.8 in AD/NC classification, and the AUC value of the other 3 classification tasks is lower than 0.8, and the AUC value of AD/NC/AMD classification is as low as 0.71, which indicates that the auxiliary diagnosis performance of ResNet-20_scratch model is not outstanding enough, and there is room for further improvement.

On the whole, the AUC values obtained by ResNet-20_scratch model in 4 classification tasks are lower than those obtained by ResNet-20_TL model, which also proves the effectiveness of the proposed method. It can be seen from the figure that the AUC of the 4 classification tasks obtained by ResNet-20_TL model is higher than 0.8, and the AUC value of AD/NC classification is up to 0.97, which shows that the ResNet-20_TL model proposed in this paper can help doctors to complete the diagnosis of AD subjects and NC subjects accurately.

CONCLUSION

With the increase of AD or AMD in the world, as the early stage of AD, the accurate diagnosis of AMD has become a hot spot in the field of scientific research. In the diagnosis research of AMD, there is still a problem that the classification accuracy is not high enough. ANN has a wide application prospect in many disciplines, such as forensic science. At present, a new upsurge of ANN research is being set off at home and abroad, and new networks with different architectures are constantly coming out, among which BP network, Hopfield network, stochastic neural network, self-organizing neural network, associative memory neural network and CMAC model are more mature. Experiments show that the EEG feature extraction method based on multispectral image fusion in multiple brain regions can obviously improve the classification accuracy of task EEG signals between AMD and NC. Based on the pre-trained ResNet-50 network, the proposed method is evaluated in ADNI data set, and the detailed experimental results are given. The results show that the classification accuracy of the proposed method is better than that of the network model trained from scratch.

In the future, we can consider the positive and negative factors that affect the relationship between brain regions when fusing multispectral images for multiple brain regions, so as to obtain multispectral images that express brain features more accurately. We can try to use 3D convolution neural network to train the above data set.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HY and MZ: conceptualization. YW: methodology and data curation. WM: software and writing—original draft preparation. HY, MZ, and WM: validation. LW: formal analysis and investigation. HY: resources, supervision, project administration, and funding acquisition. MZ: writing—review and editing and visualization. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the Natural Science Foundation of Heilongjiang Province: Effects of dihydroartemisinin on the prevention and treatment of Alzheimer's disease in mice with age-related macular degeneration (Grant No. LH2021H122) and Basic scientific research Funds of Heilongjiang Provincial undergraduate universities science and technology research project: Study on the association between Alzheimer's disease and age-related macular degeneration based on the down-regulation of Aβ6E10 metabolism by Amylin (Grant No. 2020-KYYWF-0012).

REFERENCES

1. Schwaber EJ, Thompson AC, Smilnak G, Stinnett SS, Lad EM. Co-Prevalence of Alzheimer's disease and age-related macular degeneration established by histopathologic diagnosis. *J Alzheimers Dis.* (2020) 76:207–15. doi: 10.3233/JAD-200111
2. Choi S, Sang MP, Wan JJ, Jee D. Reply to comment on: association of age-related macular degeneration on Alzheimer or Parkinson disease: a retrospective cohort study. *Am J Ophthalmol.* (2020) 213:320–1. doi: 10.1016/j.ajo.2020.01.030
3. Ong SS, Proia AD, Whitson HE, Murali SFP, Lad DEM. Ocular amyloid imaging at the crossroad of Alzheimer's disease and age-related macular degeneration: implications for diagnosis and therapy. *J Neurol.* (2019) 266:1566–77. doi: 10.1007/s00415-018-9028-z
4. Hwang DK, Hsu CC, Chang KJ, Chao D, Sun CH, Jheng YC, et al. Artificial intelligence-based decision-making for age-related macular degeneration. *Theranostics.* (2019) 9:232–45. doi: 10.7150/thno.28447
5. Liu S, Sun YP, Gao XL, SuiY. Knowledge domain and emerging trends in Alzheimer's disease:a scientometric review based on CiteSpace analysis. *Res Nerve Regen China.* (2019) 14:1643–50. doi: 10.4103/1673-5374.255995
6. Basher A, Kim BC, Lee KH, Jung HY. Volumetric feature-based Alzheimer's disease diagnosis from sMRI data using a convolutional neural network and a deep neural network. *IEEE Access.* (2021) 9:1. doi: 10.1109/ACCESS.2021.3059658
7. Fan P, Qi X, Sweet R A, Wang L. Network systems pharmacology-based mechanism study on the beneficial effects of vitamin D against psychosis in Alzheimer's disease. *Sci Rep.* (2020) 10:6136. doi: 10.1038/s41598-020-63021-8
8. Garri DD, Saakyan SV, Khoroshilova-Maslova IP, Tsygankov AY, Tarasov GY. Methods of machine learning in ophthalmology: review. *Ophthalmol Russia.* (2020) 17:20–31. doi: 10.18008/1816-5095-2020-1-20-31
9. Vellakani S, Pushbam I. An enhanced OCT image captioning system to assist ophthalmologists detecting and classifying eye diseases. *J X-Ray Sci Technol.* (2020) 28:1–14. doi: 10.3233/XST-200697
10. Heo TY, Kim KM, Min HK, Gu SM, Kim JH, Yun J, et al. Development of a deep-learning-based artificial intelligence tool for differential diagnosis

between dry and neovascular age-related macular degeneration. *Diagnostics.* (2020) 10:261. doi: 10.3390/diagnostics10050261
11. Lin AC, Lee CS, Blazes M, Lee AY, Gorin MB. Assessing the clinical utility of expanded macular OCTs using machine learning. *Transl Vis Sci Technol.* (2021) 10:32. doi: 10.1167/tvst.10.6.32
12. Kim KM, Heo TY, Kim A, Kim J, Han KJ, Min JK, et al. Development of a fundus image-based deep learning diagnostic tool for various retinal diseases. *J Pers Med.* (2021) 11:321. doi: 10.3390/jpm11050321
13. Motozawa N, An G, Takagi S, Kitahata S, Mandai M, Hirami Y, et al. Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes. *Ophthalmol Ther.* (2019) 8:527–39. doi: 10.1007/s40123-019-00207-y
14. Karthikeyan S, Sanjay KP, Madhusudan RJM, Sundaramoorthy SKS, Krishnan PK. Detection of multi-class retinal diseases using artificial intelligence: an expeditious learning using deep CNN with minimal data. *Biomed Pharmacol J.* (2019) 12:1577–86. doi: 10.13005/bpj/1788
15. Ralbovsky NM, Halámková L, Wall K, Anderson-Hanley C, Lednev IK. Screening for Alzheimer's disease using saliva: a new approach based on machine learning and raman hyperspectroscopy. *J Alzheimers Dis.* (2019) 71:1351–9. doi: 10.3233/JAD-190675
16. Bellver-Sanchis A, Pallàs M, Grián-Ferré C. The contribution of epigenetic inheritance processes on age-related cognitive decline and Alzheimer's disease. *Epigenomes.* (2021) 5:15. doi: 10.3390/epigenomes5020015
17. Noella N. Diagnosis of Alzheimer's and Parkinson's disease using artificial neural network. *Int J Scie Technol Res.* (2020) 9:3659. doi: 10.1007/978-3-642-15246-7_72
18. Duan F, Huang Z, Sun Z, Zhang Y, Sole-Casals J. Topological network analysis of early Alzheimer's disease based on resting-state EEG. *IEEE Trans Neural Syst Rehabil Eng.* (2020) 28:2164–72. doi: 10.1109/TNSRE.2020.3014951
19. Suárez-Araujo CP, García Báez P, Cabrera-León Y, Prochazka A, Rodríguez Espinosa N, Fernández Viadero C, et al. A real-time clinical decision support system, for mild cognitive impairment detection, based on a hybrid neural architecture. *Comput Math Methods Med.* (2021) 2021:5545297. doi: 10.1155/2021/5545297
20. Song X, Mao M, Qian X. Auto-Metric graph neural network based on a meta-learning strategy for the diagnosis of Alzheimer's disease. *IEEE J Biomed Health Inform.* (2021) 25:3141–52. doi: 10.1109/JBHI.2021.3053568
21. Zigunovs M. The Alzheimer's disease impact on artificial neural networks. In: *Environment Technology Resources Proceedings of the International Scientific and Practical Conference* (Rezekne). (2021). p. 205–9. doi: 10.17770/etr2021vol2.6632
22. Swietlik D, Białowas J. Application of artificial neural networks to identify Alzheimer's disease using cerebral perfusion SPECT data. *Int J Environ Res Public Health.* (2019) 16:1303. doi: 10.3390/ijerph16071303
23. Tan HD, Lv M, Tan X, Su GN. Sharing of genetic association signals by age-related macular degeneration and Alzheimer's disease at multiple levels. *Mol Neurobiol.* (2020) 57:4488–99. doi: 10.1007/s12035-020-02024-y
24. Kun C, Feiyi S, Jian D, Chen F, Han X. Network pharmacology-based study on the mechanism of *Schisandra chinensis* for treating Alzheimer's disease. *Indian J Pharmacol.* (2020) 52:94–101. doi: 10.4103/ijp.IJP_515_19
25. Sharunya S R, Desai V, Singh M, Kusuma M. Survey on early detection of Alzhiemer's disease using capsule neural network. *Int J Art Intell.* (2020) 7:7–12. doi: 10.36079/lamintang.ijai-0701.65
26. Jo T, Nho K, Risacher SL, Tan D. The deep learning of neuroimaging, genetics and demographic data is used to diagnose and classify Alzheimer's disease. *Alzheimers Dis Dement.* (2019) 15:P409–10. doi: 10.1016/j.jalz.2019.06.1003
27. Akinori M, Ryozo K, Takeshi I, Iijima KM, Akihiro N. Disruption of a RAC1-centred network is associated with Alzheimer's disease pathology and causes age-dependent neurodegeneration. *Hum Mol Genet.* (2020) 29:817–33. doi: 10.1093/hmg/ddz320
28. Kumar S, Ansari MD, Gunjan VK, Solanki VK. On classification of BMD images using machine learning (ANN) algorithm. In: *ICDSMLA 2019.* Singapore: Springer (2020). p. 1590–9. doi: 10.1007/978-981-15-1420-3_165
29. Rashid E, Ansari MD, Gunjan VK, Khan M. Enhancement in teaching quality methodology by predicting attendance using machine learning technique. In: *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough.* Cham: Springer (2020). p. 227–35. doi: 10.1007/978-3-030-38445-6_17
30. Xin H, Shuqin L, Bin L. Identification of grape leaf diseases based on multi-scale residual neural network. *J Comput Engg.* (2021) 47:285–91. doi: 10.19678/j.issn.1000-3428.0057818

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Gong, Ma, Wen, Wang and Yao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 04 July 2022
doi: 10.3389/fpubh.2022.924432

Breast Cancer Detection and Classification Empowered With Transfer Learning

Sahar Arooj[1], *Atta-ur-Rahman*[2], *Muhammad Zubair*[3], *Muhammad Farhan Khan*[4], *Khalid Alissa*[5], *Muhammad Adnan Khan*[6]* *and Amir Mosavi*[7,8,9]

[1] *Riphah School of Computing and Innovation, Riphah International University Lahore, Lahore, Pakistan,* [2] *Department of Computer Science, College of Computer Science and Information Technology (CCSIT), Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia,* [3] *Faculty of Computing, Riphah International University Islamabad, Islamabad, Pakistan,* [4] *Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan,* [5] *Networks and Communications Department, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia,* [6] *Department of Software, Gachon University, Seongnam, South Korea,* [7] *John von Neumann Faculty of Informatics, Obuda University, Budapest, Hungary,* [8] *Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology in Bratislava, Bratislava, Slovakia,* [9] *Faculty of Civil Engineering, TU-Dresden, Dresden, Germany*

OPEN ACCESS

Edited by:
Ali Kashif Bashir, Manchester Metropolitan University, United Kingdom

Reviewed by:
Noor Zaman Jhanjhi, Taylor's University, Malaysia
Dilbag Singh, Gwangju Institute of Science and Technology, South Korea

***Correspondence:**
Muhammad Adnan Khan
adnan@gachon.ac.kr

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *20 April 2022*
Accepted: *31 May 2022*
Published: *04 July 2022*

Citation:
Arooj S, Atta-ur-Rahman, Zubair M, Khan MF, Alissa K, Khan MA and Mosavi A (2022) Breast Cancer Detection and Classification Empowered With Transfer Learning. Front. Public Health 10:924432. doi: 10.3389/fpubh.2022.924432

Cancer is a major public health issue in the modern world. Breast cancer is a type of cancer that starts in the breast and spreads to other parts of the body. One of the most common types of cancer that kill women is breast cancer. When cells become uncontrollably large, cancer develops. There are various types of breast cancer. The proposed model discussed benign and malignant breast cancer. In computer-aided diagnosis systems, the identification and classification of breast cancer using histopathology and ultrasound images are critical steps. Investigators have demonstrated the ability to automate the initial level identification and classification of the tumor throughout the last few decades. Breast cancer can be detected early, allowing patients to obtain proper therapy and thereby increase their chances of survival. Deep learning (DL), machine learning (ML), and transfer learning (TL) techniques are used to solve many medical issues. There are several scientific studies in the previous literature on the categorization and identification of cancer tumors using various types of models but with some limitations. However, research is hampered by the lack of a dataset. The proposed methodology is created to help with the automatic identification and diagnosis of breast cancer. Our main contribution is that the proposed model used the transfer learning technique on three datasets, A, B, C, and A2, A2 is the dataset A with two classes. In this study, ultrasound images and histopathology images are used. The model used in this work is a customized CNN-AlexNet, which was trained according to the requirements of the datasets. This is also one of the contributions of this work. The results have shown that the proposed system empowered with transfer learning achieved the highest accuracy than the existing models on datasets A, B, C, and A2.

Keywords: breast cancer (BC), deep learning (DL), learning rate (LR), machine learning (ML), transfer learning (TL), convolutional neural network (CNN)

INTRODUCTION

Medical imaging is a valuable tool for detecting the existence of various medical diseases and analyzing investigational outcomes. The use of biomedical imaging in cancer treatment is crucial. Cancer is a major public health issue in the modern world. According to the World Health Organization (WHO), cancer in 2018 caused 9.6 million deaths, and a probable 10 million deaths were caused by cancer in 2020 (1). Cancer tumors are caused by the uncontrollable growth of cells in the breast. One of the most frequent malignancies in women is breast cancer. BC is estimated to attack more than 8% of women at some point in their life. BC can start in any part of the breast. The majority of BC begins in the lobules or ducts. However, BC can be detected early, allowing patients to obtain proper therapy and so increase their chances of survival.

Imaging technologies such as magnetic resonance imaging (MRI), diagnostic mammography (2) (X-rays), thermography, and ultrasound (sonography) can help analyze and identify breast cancer (3). Ultrasound images are used in this proposed study. Breast cancer is classified as benign and malignant. Benign tumor cells only grow in the breast and do not split throughout the other cells. A malignant tumor is made up of cancerous cells that have the ability to expand uncontrollably, spread to other areas of the body, and infect other tissues. Because cancer cells vary in size, shape, and location, automatically detecting and localizing cancer cells in BC images are a huge difficulty. Machine learning (ML) (4) approaches have found widespread use in a variety of domains, including educational prediction, pattern recognition, image editing, feature reduction, defect diagnosis, face identification, micro-expression recognition, NLP, and medical diagnosis. Its greatest potential has been discovered in the diagnosis of breast cancer (5).

Many researchers have proposed numerous strategies for the automatic classification of cells in breast cancer detection in recent decades (6). By identifying nucleus traits, cancerous cells of breast cancer can be classified as benign and malignant. However, the system's efficiency and accuracy decrease as a result of the complexity of typical machine learning procedures such as pre-processing, segmentation, feature extraction, and others. Traditional ML problems can be solved using the recently developed DL technique. With exceptional feature representation, this technique can perform picture classification and object localization challenges. The transfer learning approach used a natural-image dataset such as ImageNet and then applied a fine-tuning technique to solve this problem. The main benefit of transfer learning is that it improves classification accuracy and speeds up the training process.

First, network parameters were pre-trained using the data and used in the required domain, and then the system restrictions were changed for improved performance. This study used a model for the classification and detection using TL. The proposed model has two components. The first component is training, and the second component is testing. BC classification can be done using a CNN pre-trained such as the ResNet50, VGG 16, VGG 19, and Inception V2 Res Net. In this work, we have done the job of BC classification and detection by using the AlexNet model. AlexNet is a powerful model that can achieve high accuracies on even the most difficult datasets. AlexNet is a leading architecture for any object identification task and classification, and it has a wide range of applications in the artificial intelligence field of computer vision. Some previous studies used the AlexNet, but in this work, we used a customized AlexNet model which has not been used before in previous studies. In the customized AlexNet, the first and last three layers of the architecture are modified, and newly modified layers are the image input layer, fully connected layer, classification layer, and softmax layer, although the remaining layers remain fixed. The customized model has all of the features for image processing that it learned during the process of training. The main goal of this project was to detect and classify breast cancer, reduce training time, increase accuracy, and enhance classification performance.

There are many previous studies on breast tumors using various types of models, but with some limitations, breast cancer has limited studies due to the lack of publicly available benchmark datasets. This proposed system worked on three datasets A, B, C, and A2, A2 is dataset A with two classes with a total number of 10,336 images, which is a good dataset. This study is the first to compare three common datasets and suggest the use of customized transfer learning algorithms for breast cancer classification and detection on multiple datasets. By using the customized AlexNet, we achieved the optimum accuracy. This work used ultrasound images and histopathology images, the sample images of ultrasounds are shown in **Figure 1**, and the sample images of histopathology are shown in **Figure 2**.

This study is divided into five sections. Section 2 is the literature review, section 3 is the proposed system model, section 4 is the simulation and results, and section 5 is the conclusion of this work.

LITERATURE REVIEW

Diagnosis of BC disease is a challenge for researchers. To solve this problem of breast cancer, various models and techniques such as ML, DL, and TL are used. Researchers used datasets based on mammography (X-rays), magnetic resonance imaging (MRI), ultrasound (sonography), and thermography to diagnose breast cancer disease.

Fractal dimension (FD) is the best indicator of ruggedness for regular elements, according to their findings. Breast lumps are uneven and can vary from malignant to benign; as a result, the breast is one of the best places to apply fractal geometry. The support vector machine, on the other side, is a new categorization technique. They (2) employed two techniques, FA: SVM and Box Count Method (BCM) in distinct operations that produced good results in respective sectors. The BCM is used to extract features. The retrieved feature "FD" assesses the difficulty of the 42-image input dataset. The generated FD is then processed using the SVM classifier which is used to classify malignant and benign cells. Their highest accuracy is 98.13%.

Breast cancer is a major disease among women between the ages of 59 and 69. They (4) also showed that finding tiny tumors early improves predictions and reduces death rates

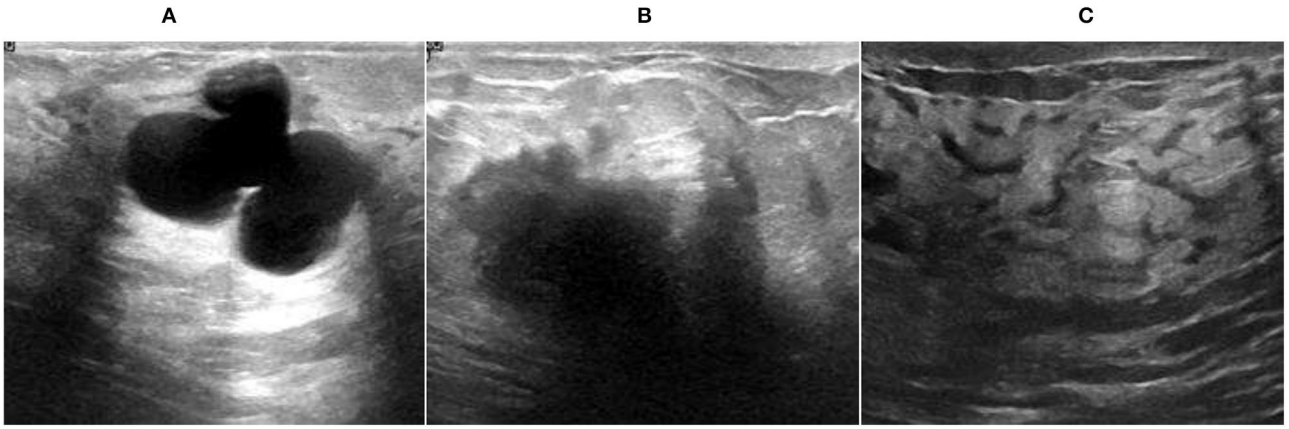


FIGURE 1 | Ultrasound image samples: **(A)** benign, **(B)** malignant, and **(C)** normal.

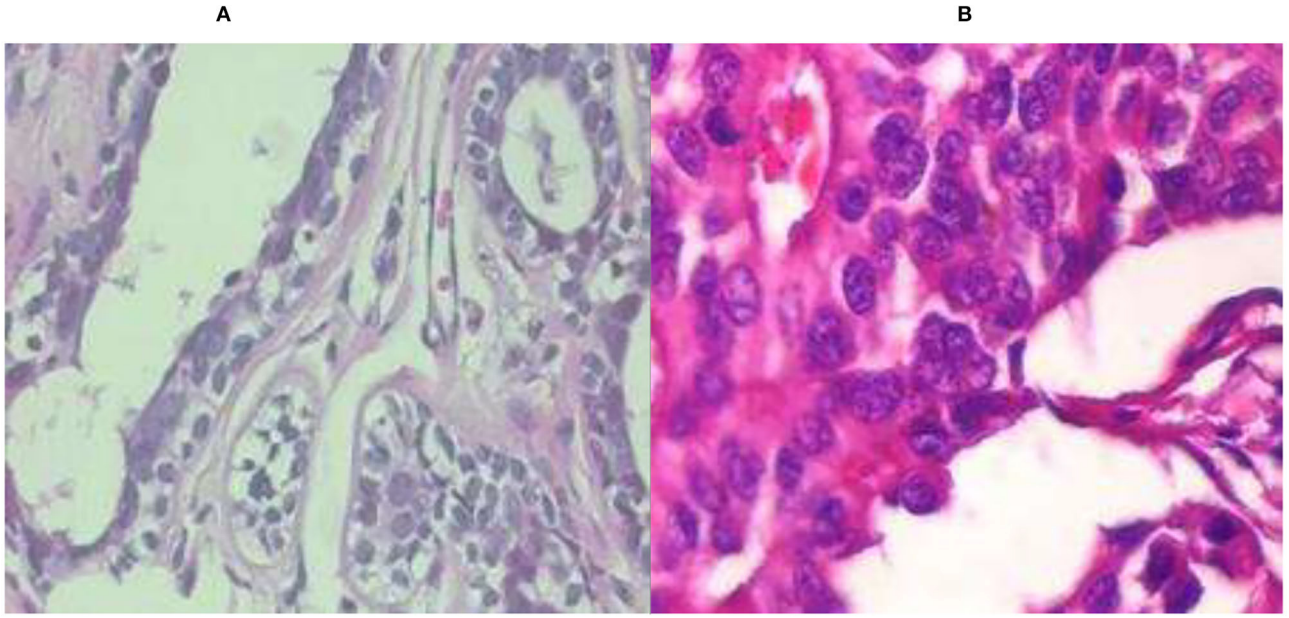


FIGURE 2 | Histopathology image samples: **(A)** benign and **(B)** malignant.

significantly. Mammography is a useful screening diagnostic method. However, due to tiny changes in tissue densities within mammography pictures, mammography interpretation is challenging. This is particularly true for solid tissues of the breast, and according to this study, screening is more appropriate in greasy breast tissue than in solid breast tissue. Their research focuses on BC detection, as well as danger issues and breast cancer assessments. Their research also focuses on the early diagnosis of BC using 3D MRI mammographic technologies and the classification of mammography pictures using ML.

Their research (5) proposes a heterogeneous efficient machine learning strategy for the early detection of breast cancer. The suggested method follows the CRISP-DM process and employs a stack to construct the collaborative model, which involves three algorithms: KNN, SVM, and decision tree. This meta-classifier's performance is compared to the separate presentations of DT, SVM, and KNN and other particular classifiers NB, SGD, LR, ANN, and a homogeneous collaborative model of (KNN, SVM, DT) and (RF). Using chi-square, the top five characteristics such as glucose, resist in, HOMA, insulin, and BMI are calculated. At K = 20, the proposed collaborative model has the best accurateness of 78% and the smallest log loss of 0.56, denying the null hypothesis. The one-tailed *t*-test, which delivers a lesser consequence at $\infty = 0.05$, yields a *P*-value of 0.014.

In this paper (7), they tested the presentation of using conveyed features from a pre-trained model on a dataset of

1,125 breast ultrasound cases. Their dataset is composed of 2,392 regions of interest (ROIs). Each ROI was marked up as cystic, malignant, or benign. Using a convolutional neural network (CNN) (6) from each ROI, features were taken out and used to train (SVM) classifiers. For comparison, classifiers were also trained before retrieving tumor features. CNN-extracted feature-trained classifiers were pretty similar to human-designed feature-trained classifiers. The SVM (8) which was trained on both human-designed features and CNN-extracted features had a 90% accuracy rate in the classification task. The accuracy of the SVM trained on CNN features was 88%, compared to 85% for the SVM trained on features that are human-designed in the task of determining malignant or benign. Deep learning (DL) methods currently in use rely on large datasets. It is worth noting that the study's dataset is not available to the general public.

In this work (8), they look at the potential uses of machine learning for brain problems. They show why machine learning is generating so much interest among researchers and clinicians in the field of brain disorders (9) by highlighting three main applications: predicting sickness onset, assisting with diagnosis, and predicting longitudinal outcomes. They explore the hurdles that must be solved for a successful translational implementation of machine learning in routine psychiatric and neurologic care after exhibiting various applications.

This paper (10) used two datasets of breast ultrasound from two different systems. The first set of data is called breast ultrasound images (BUSI). There is a total of 780 photographs in the BUSI dataset (normal 133, malignant 210, and 437 benign). B dataset has 163 pictures (110 benign and 53 malignant). They used a generative adversarial network (GAN) technology for data augmentation. Researchers can access their BUSI dataset for free. In addition, DL algorithms are applied in this study for breast ultrasound classification. They compare the performance of two alternative methods: a CNN-AlexNet approach and a transfer learning technique with and without augmenting. Their network is trained with a 0.0001 learning rate and 60 epochs. They achieved the accuracies of 94% on BUSI data, 92% on dataset B, and 99% on augmentation.

In this paper (11), they introduce a publicly available collection of 7,909 breast cancer histopathology images. Both benign and malignant images are included in the dataset. The aim connected with this dataset is to automatically classify these photographs into two categories which would be a useful computer-aided diagnosis tool for the clinician. The accuracy ranges from 80 to 85% indicating that there is still space for improvement. In their work to evaluate the feature collection, they used multi-classifiers KNN, SVM, quadratic linear analysis, and RF.

The use of DL techniques for breast ultrasound lesion identification is proposed in this study (12), and three alternative methods are investigated: patch-based Le Net, transfer learning (13), and U-Net approach with the AlexNet model. Two conventional ultrasound picture datasets were obtained, and two separate ultrasound devices are compared and contrasted in this study. Dataset A contains 306 photographs (246 benign and 60 malignant), while dataset B has a total of 163 images (110 benign and 53 malignant). They employed grayscale ultrasound pictures that were divided into 28 × 28 patches. RMS Propagation with LR of 0.01 and 60 epochs is used to train the network. They used the AlexNet model to attain a maximum accuracy which is 91% for dataset A and 89% for dataset B.

Based on two methodologies cross-validation and 80–20, a DL model based on the TL methodology is built in this study (14) to proficiently help in the automatic identification and identification of the breast cancer suspicious area. Deep learning architectures are designed to solve certain problems. Transfer learning applies what one has learned while working on one problem to another. They used six evaluation metrics to assess the proposed model's performance. To train this model, they used a learning rate of 0.01 and 60 epochs. Transfer learning is effective in detecting breast cancer by categorizing mammogram images of the breast with general accuracy, sensitivity, specificity, precision, F-score, and accuracy of 98.96, 97.83, 99.13, 97.35, 97.6.%, and 95%, respectively.

They (15) investigate a quantitative solution to a machine learning problem in this paper. They used transfer learning to train a set of hybrid traditional neural networks based on Azevedo et al. (15) work. Their mission was to tackle BCDR's difficulty in identifying full-image mammograms as malignant or benign. Data collected in this study were used throughout our research to illustrate the regions of the mammograms that the networks were targeting while measuring various performance indicators. They also indicate that some designs perform much better than others depending on the task. According to their findings, the greatest accuracy is 84%.

They (16) demonstrate in their study that the early detection and classification of breast cancer are critical in assisting patients in taking appropriate action. Mammography images, on the contrary, have low sensitivity and efficiency for identifying breast cancer. Furthermore, MRI has a higher sensitivity for detecting breast cancer than mammography. A novel Back Propagation Boosting Recurrent Widening Model (BPBRW) with a Hybrid Krill Herd African Buffalo Optimization (HKH-ABO) method is created in this study to diagnose breast cancer at an earlier stage utilizing breast MRI data. The system is initially trained using MRI breast pictures. Furthermore, the proposed BPBRW with HKH-ABO mechanism distinguishes between benign and malignant breast cancer tumors. Additionally, Python is used to simulate this model. They demonstrate that their model has a 99.6% accuracy rate.

They (17) constructed four distinct predictive models and offered data exploratory techniques (DET) to increase breast cancer detection accuracy in this study. Prior to the models, researchers dug deep into four-layered critical DET, such as feature distribution, correlation, removal, and hyperparameter optimization, to find the most robust feature categorization into malignant and benign classifications. On the WDBC and BCCD datasets, the proposed approaches and classifiers were tested. To evaluate each classifier's efficiency and training time, standard performance metrics such as confusion matrices and K-fold approaches were used. With DET, the models' diagnostic capacity improved, on polynomial SVM achieving 99.3% accuracy, LR 98.06, KNN 97.35, and EC 97.61% accurateness with the WDBC database.

Their (18) goal was to create a hierarchical breast cancer system model that would improve detection accuracy and reduce breast cancer misdiagnosis. To categorize breast cancer tumors and compare their performances, the dataset was subjected to ANN and SVM. The SVM utilizing radial features produced the best accuracy of classification of 91.6%, whereas the ANN obtained 76.6%. As a result, SVM was used to conclude about the importance of breast screening. The second stage involved applying transfer learning to train AlexNet, InceptionV3, and ResNet101. AlexNet scored 81.16%, ResNet101 scored 85.51%, and InceptionV3 scored 91.3 %, according to the data.

They (19) present a framework based on the notion of transfer learning in their research. In addition, a variety of augmentation procedures, including multiple rotation combinations, scale, and shifting, were implemented to prevent a fitting problem and create consistent outcomes by expanding the number of screened mammography pictures. Their proposed solution was tested on the Screening mammography Image Analysis Society (MIAS) database and achieved an accuracy of 89.5% using ResNet50 and 70% utilizing the NASNet-Mobile network. Pre-trained categorization networks are much more efficient and effective, making them more suitable for diagnostic imaging, especially for short training datasets, according to their suggested system.

They (20) used machine learning-based algorithms to help the radiologist read mammography pictures and classify the tumor in an acceptable amount of time in this study. They extracted a number of features from the mammogram's region of interest, which the physician manually labeled. To train and create the suggested structural classification models, these properties are added to a classification engine. They tested the suggested system's accuracy using a dataset that had never been encountered before in the model. As a result, this research discovered that a variety of circumstances can affect the results, which they ignored after investigating. After merging the selection of features optimization approaches, this study advises employing the optimized SVM or Nave Bayes, which provided 100% accuracy.

Their (21) research focuses on employing TL with fine-tuning and on training the CNN with areas derived from the IN breast and MIAS datasets to apply, evaluate, and compare architectures such as AlexNet, Google Net, Vgg19, and Resnet50 to classify breast lesions. They looked at 14 classifiers, each of which corresponded to benign or malignant microcalcifications and masses, as several previous studies have done. With the CNN, they obtained the best results. With an AUC of 99.29%, an F1 score of 91.92%, accuracy of 91.92%, precision of 92.15%, sensitivity of 91.70%, and specificity of 97.66% on a balanced database, Google Net is the better model in a Cad model for breast cancer.

The effectiveness of BC categorization for malignant and benign tumors was evaluated utilizing several machine learning algorithms (k-NN, RF, and SVM) and aggregation methods to calculate the prediction of BC survival by applying 10-fold cross-validation. Their research (22) used a dataset from WDBC that included 23 selected variables evaluated by 569 people, of whom 212 had malignant tumors and 357 had benign tumors. The analysis was done to look at the characteristics of the tumors using the mean, worst values, and standard error. There are 10 properties for each feature. According to the results, AdaBoost has the maximum accuracy for 30 features (98.95%), 10 mean features (98.07%), and 10 worst features (98.77%) with the lowest error rate. To obtain the best accuracy rate, their recommended approaches are categorized using 2-, 3-, and 5-fold cross-validation. When all approaches were compared, AdaBoost ensemble methods had the highest accuracy, with 98.77% for 10-fold cross-validation and 98.41 and 98.24% for 2- and 3-fold cross-validation, respectively. Nonetheless, 5-fold cross-validation revealed that SVM generated the highest accuracy rate of 98.60% with the least error rate.

Breast cancer affects a large number of people all around the world. Mammography is a key advancement in breast cancer detection. It is difficult for doctors to recognize due to its intricate structure. Their (23) research suggests using a CNN to detect cancer cells early. By separating malignant and benign mammography pictures, detection and accuracy can be greatly improved. The Break His $\times$400 database comes from Kaggle, and the architectures NASNet-Large, DenseNet-201, Big Transfer (M-r101x1x1), and Inception ResNet-V3 perform admirably. M-r101x1x1 has a maximum accuracy of 90% among them. The most important goal of their research is to use selected neural networks to accurately classify breast cancer. This research could help to enhance the systematic diagnosis of early-stage breast cancer.

Despite the fact that there are several scientific studies on the categorization and identification of cancer tumors using various types of models but with some limitations. Breast cancer has limited studies due to the lack of publicly available benchmark datasets. In their work (14), they have used multiple methods such as ResNet50, inception V3, Inception V2 Res Net VGG 19, and VGG 16 but their dataset is too small and they just work on one single dataset and their maximum accuracy is 98.96. In this work (10), they used two different datasets using transfer learning. Datasets are good, but their maximum accuracy is 94% on the BUSI dataset and 92% on dataset B. In this work (12), they also used two different datasets by using CNN multiple models, and they achieved a maximum accuracy with AlexNet, 91% on dataset A and 89% on dataset B. In their work (11), they used a good and large dataset but they also achieved a maximum accuracy is 80–85%. **Table 1** shows the comparison of previous studies in terms of accuracy and limitations. Previous studies (4, 5, 7, 10–12, 14) have some limitations like less number of images in the dataset, less accuracy, hand-crafted features required, lack of diverse datasets, no publically available dataset, and an imbalanced number of images in datasets.

The following are the primary contributions of this work:.

- This work used three different datasets of breast cancer and compare their results on the same model.
- Improving the accuracy of classification and detection by customizing the model AlexNet.
- Model proposed achieved the maximum accuracy results using transfer learning approaches.

TABLE 1 | Comparison and limitations of previous studies.

Publication	Model	Accuracy	Dataset	Limitations
Swain et al. (4)	SVM	98.13%	Private	• Dataset is small
Nanglia et al. (5)	KNN+ SVM + DT	78%	Private	• Less accuracy • Required hand crafted features
Krizhevsky et al. (7)	SVM	88%	Private	• Required hand crafted features • Dataset is not available publically
Dhabyani, et al. (10)	AlexNett+ VGG 16+ Inception+ Res net+ Nasnet	78%, 88%, 85%, 93%, 94%, 80%, 82%, 80%, 90%, 92%	Public	• Less accuracy
Spanhol et al. (11)	SVM	80%	Private	• Less accuracy • Required hand-crafted features
Yap et al. (12)	AlexNet	91%, 89%	Private	• Use of imbalanced dataset • Dataset is small
Saber et al. (14)	Inception v3 + Resnet 50+ VGG 16+ VGG 19	96%, 94%, 96%, 95%	Public	• Dataset is small • Lack of diverse dataset

PROPOSED MODEL

In order to assess and identify diseases in medical images, machine learning techniques were applied. Many ML (24) and DL (25) approaches have been widely employed in medical image processing in recent years to detect and evaluate items in medical images. The use of DL techniques to detect breast cancer at an early stage aids medical practitioners in determining its therapy. Breast cancer has been diagnosed early using a variety of DL and transfer learning approaches. DL methods are useful tools for detecting the disease early. **Figure 3** shows the application-level representation of the suggested system model.

As we know that in deep learning, few steps are very important: first is data acquisition, data pre-processing, and then the training of the datasets. As we know that if the data are image-based, then the deep learning methods give more accurate results as compared to the machine learning, which is the reason we used deep learning-based solution. There are various kinds of deep learning models like CNN and KNN, as we know that computational resources are also required to compute such kinds of problems like processing power. In further deep learning if we have less computational resources like this one, then we used transfer learning instead of the other deep learning models that is why here we used transfer learning to save the computing power resource optimization. In transfer learning, we used a pre-trained model AlexNet, and we customized this model according to our problem which saves computing power. After that, we stored it in the cloud so that we can use this pre-trained model.

The detailed proposed system model is shown in **Figure 4**. The projected method for breast cancer identification and classifications contains two major components. The first component is pre-processing and training, and the second is testing. Based on deep learning techniques, the proposed system model accepts images to help in the classification and early detection of diseases in various stages. Previous research and the Kaggle repository were used to collect the training data, which consisted of ultrasound and histopathology images and the data were collected in raw form. The raw data were handled by the pre-processing layer, which converted the images according to the requirement of the model which is 227*227 for AlexNet and customized the pre-trained model AlexNet for transfer learning.

The second layer is training, and for training, this study used pre-processed images of the training layer [227*227] and import the customized trained model. The model must be retrained if the learning conditions are not met; otherwise, the trained model is saved in the cloud. The intelligent trained model detects and classifies breast cancer into three categories: benign, malignant, and normal. If the patient is normal, no need to visit the doctor, and if the patient is having symptoms of benign or malignant system, refer her to the doctor for the treatment of BC. **Table 2** shows the pseudocode of the proposed algorithm.

Dataset

In general, a dataset should be provided to construct a healthcare system employing deep learning. Three separate datasets of breast images are used in this investigation. This study referred to datasets as datasets A, B, and C. Dataset A is collected from (10, 26), dataset B is collected from (11, 27), and dataset C is collected from (28). Dataset A includes medical images of breast cancer obtained by an ultrasound scan. The images in this dataset A are divided into three categories: normal, benign, and malignant. Dataset B contains histopathology pictures of malignant and benign breast cancers, and images were taken as part of a clinical investigation. Dataset C images are also histopathology images. Dataset C is divided into two categories: malignant and benign. We also used dataset A with two classes, benign and malignant, and called it as dataset A2. The number of images in all datasets is shown in **Table 3**.

After data collection, pre-processing of images is done. This pre-processing is critical for removing the limitations of abnormality observation and dimension of images according to the AlexNet model. The quality of the images can be increased, and the results can become more precise. Splitting is an important part of a model for training and testing. This proposed model is done by splitting datasets randomly into 80% for the training set and 20% for testing.

Transfer Learning

Transfer learning is a technique that involves training a CNN model to learn features for a wide range of domains. The proposed TL method is based on AlexNet. The images of the breasts are in grayscale. To make model training easier, pre-processing actions like resizing images into 227*227 were taken. This study divided the dataset into training and testing groups randomly, so the models were able to identify significant elements in each image and get a perfect score on the test set. The

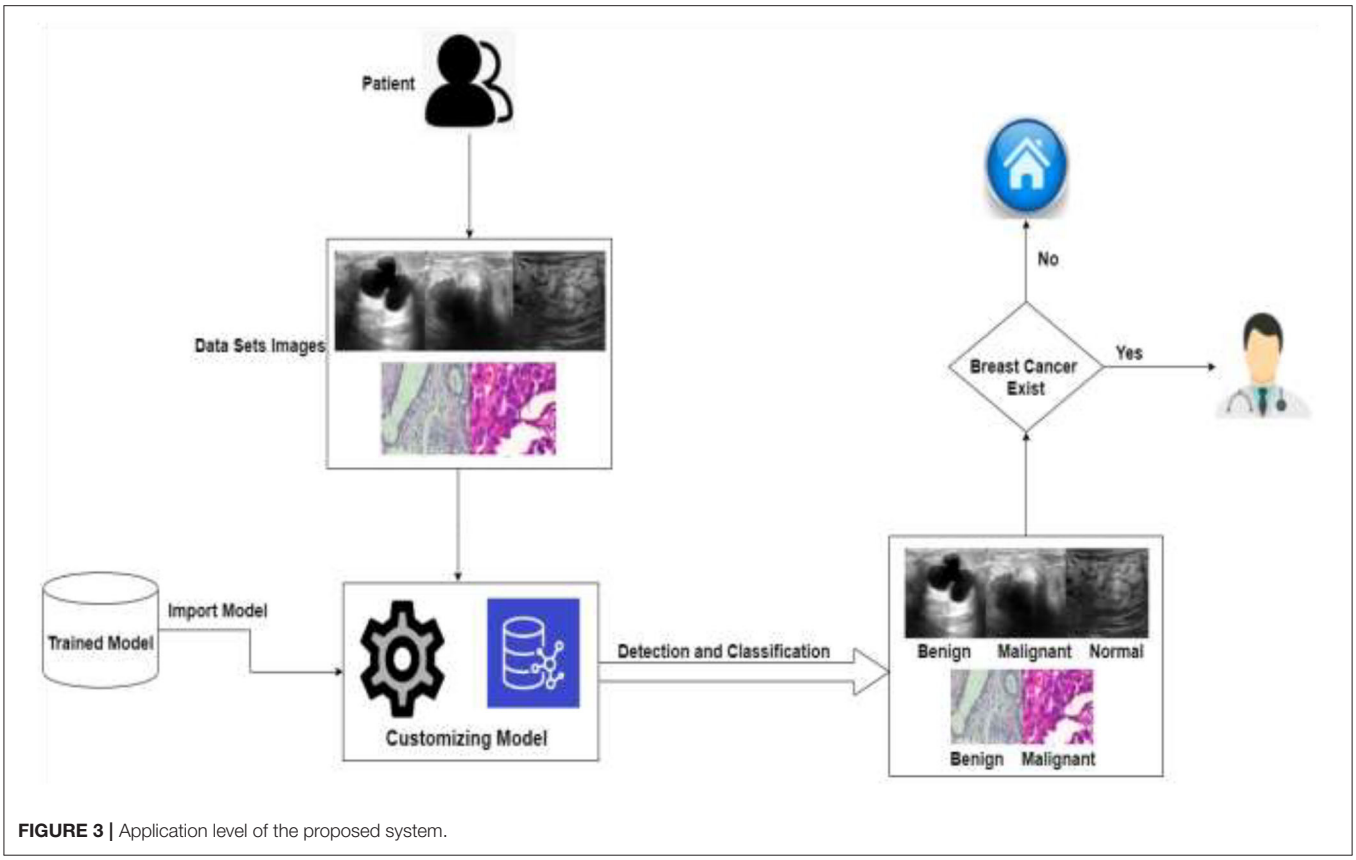


FIGURE 3 | Application level of the proposed system.

AlexNet model was used to train all datasets (A, B, and C). The model that has been trained is kept and reused.

This proposed methodology customized the AlexNet model. AlexNet is an eight-layer network with learnable parameters in which three are fully connected layers and five are convolutional layers with max pooling. ReLU is a non-linear initiation function that exists in each layer. Images from the pre-processed layer are read by the network's input layer. The fully connected layers learn disease features to categorize images into specific classes. Early convolutional layers extract common features from pictures by using filters such as detection of edges and preserving the spatial connection between pixels, but later convolutional layers using filters extract general features from images such as detection of edges.

This CNN (29) network was modified to our needs, and the pre-processed images were then loaded into the proposed AlexNet transfer learning model (30). According to the problem, the first and last three layers of the architecture are modified and newly modified layers are the image input layer, fully connected layer, softmax layer, and classification layer, although the remaining layers remain fixed. This customized network is used for TL. The first layer will set the dimension into 227*227, and the last three layers are set up according to the labels of the output class and they can categorize the images into their respective groups. The output's size, which is divided into numerous types, is the input parameter for fully connected layers. The fully connected layer in the proposed model will connect three classes: benign, malignant, and normal. Softmax layers are used to apply softmax functions to the input. A fully connected layer learns the class's precise features to differentiate across classes. So, fully connected layers are altered according to dataset classes. To identify images in distinct class labels, this projected network is trained on breast cancer labels of multi-class.

Learning rate and number of epochs are two of the parameters that can be used as training options. The learning rate and epochs are used to train the network. The training was done on various epochs such as 10, 30, and 50, and it was discovered that the ideal epoch was 50, with a learning rate of 0.0008. For training, the stochastic gradient descent with momentum (SGDM) technique of optimization is used. Newly edited layers use these training settings for the breast cancer dataset. The CNN layers are accountable for extracting the general features of images and then for the classification and identification of new datasets by reusing these learning parameters. Models that have been customized and trained are placed on clouds and can be reused. Pre-processed images are passed to the customized model AlexNet during the validation stage. The customized model has all of the features for image processing that it learned during the process of training, so it assesses the images and classifies them into normal, benign, and malignant diseases. After the classification and detection of breast cancer if the patient is normal, no need to visit a doctor, and if the patient has symptoms of disease, then refer to a doctor.

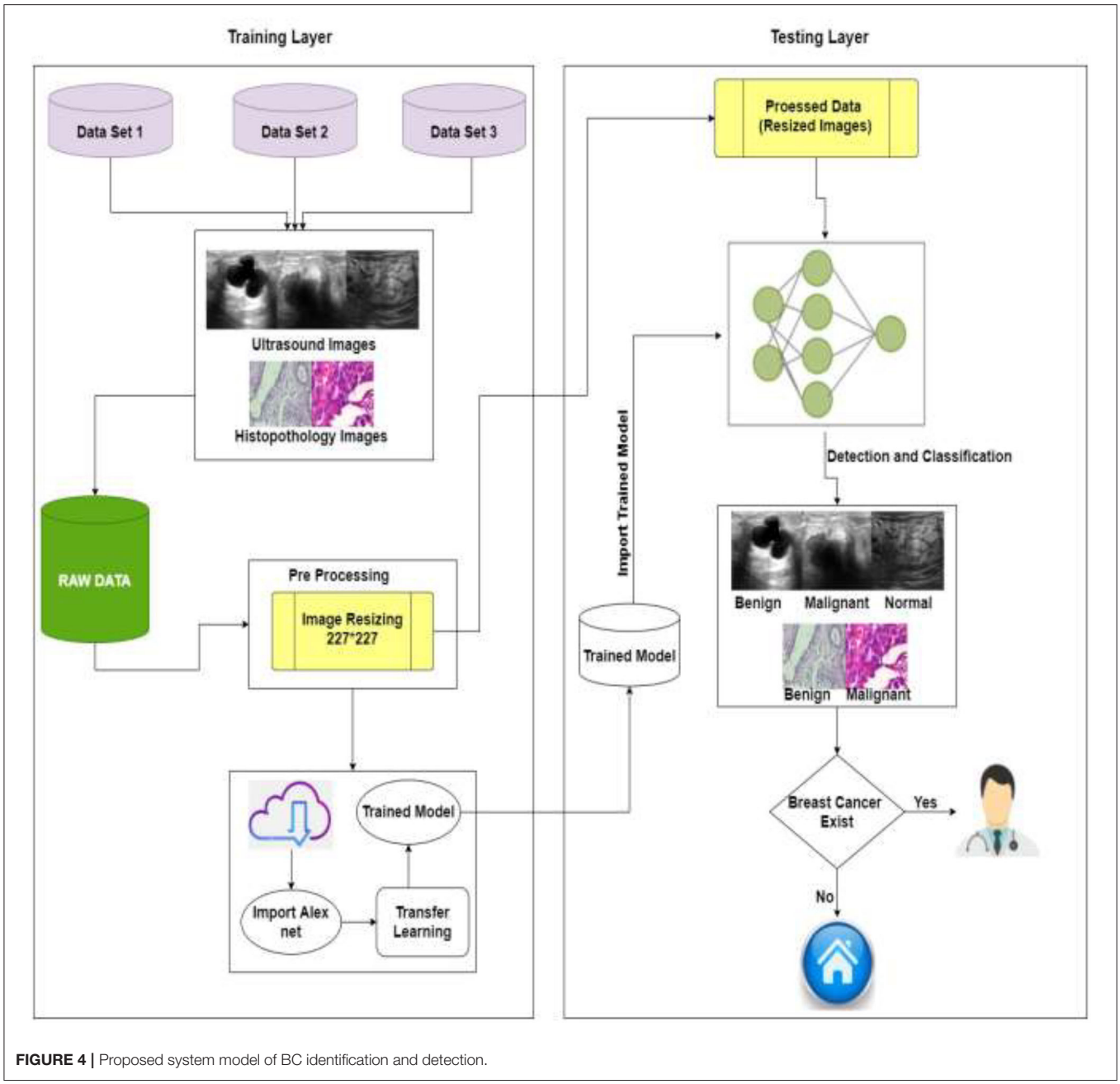


FIGURE 4 | Proposed system model of BC identification and detection.

SIMULATION AND RESULTS

Breast cancer is caused by the uncontrollable growth of cells in the breast. One of the most frequent malignancies in women is breast cancer. BC is estimated to attack more than 8% of women at some point in their life. However, BC can be detected early, allowing patients to obtain proper therapy and so increase their chances of survival. There are many previous studies on breast tumors using various types of models, but with some limitations, breast cancer has limited studies due to the lack of publicly available benchmark datasets. In this study, we worked on three datasets (A, B, C, and A2, A2 is dataset A with two classes) with a total number of 10,336 images which is a good dataset. This study is the first to compare three common datasets and suggests the use of customized transfer learning algorithm AlexNet for breast classification and detection on multiple datasets. **Table 14** shows that previous studies do not give accurate results that is why we need a better solution to diagnose breast cancer with more accuracy. Some previous studies used AlexNet, but in this work, we used customized AlexNet model that is not used before in previous studies. The customized model has all of the features for image processing that it learned during the process of training. By using customized AlexNet, we achieved good results that are shown in this section of this study.

TABLE 2 | Pseudocode of the proposed model.

1	Start
2	Input breast cancer datasets A, B, C, and A2
3	Pre-processing of the datasets
4	Load data
5	Load pre-trained model
6	Detection and classification of BC using transfer learning model (Customized Alex net)
7	Training phase
8	Store on cloud
9	Image validation phase
10	Compute the performance and accuracy of the proposed model on all dataset by using evaluation matrix
11	Finish

TABLE 3 | Dataset parameters.

Data set	No. of images
Dataset A	780 (Benign = 437, Malignant = 210, Normal = 133)
Dataset B	7,783 (Benign = 2,479, Malignant = 5,304)
Dataset C	1,126 (Benign = 547, Malignant = 579)
Dataset A2	647 (Benign = 437, Malignant = 210)

TABLE 4 | Training model on 50 epochs class-wise.

Data set	Classes	Epochs	Accuracy	Miss rate
A	Benign	50	98.9%	1.1%
	Malignant	50	100%	0.0%
	Normal	50	100%	0.0%
B	Benign	50	96.0%	4.0%
	Malignant	50	97.0%	3.0%
C	Benign	50	99.1%	0.9%
	Malignant	50	99.1%	0.9%
A2	Benign	50	100%	0.0%
	Malignant	50	100%	0.0%

TABLE 5 | Training model on dataset A.

	Epochs	Accuracy	Miss rate
Dataset A	10	70.5%	29.5%
	30	96.8%	3.2%
	50	99.4%	0.6%

TABLE 6 | Training model on dataset B.

	Epochs	Accuracy	Miss rate
Dataset B	10	77.5%	22.5%
	30	95.6%	4.4%
	50	96.7%	3.3%

TABLE 7 | Training model on dataset C.

	Epochs	Accuracy	Miss rate
Dataset C	10	96.0%	4.0%
	30	97.3%	2.7%
	50	99.1%	0.9%

TABLE 8 | Training model on dataset A2.

	Epochs	Accuracy	Miss rate
Data set A2	10	89.1%	10.9%
	30	96.1%	3.9%
	50	100%	0.0%

In this section, multiple tests were carried out to investigate the performance of this model on datasets A, B, and C. Benign, malignant, and normal were used to categorize the datasets. AlexNet (31) was used to create the proposed model for detecting and classifying breast cancer. The categorization and findings are done in MATLAB 2020a. Evaluation metrics are used to assess produced results. In the training phase, for training, 80% of the dataset is utilized while for testing 20% is used. Transfer learning is applied on AlexNet and compared in form of accuracy (Acc), sensitivity (Sen), specificity (Spe), false-negative ratio (FNR), Miss classification rate (MCR), false-positive ratio (FPR), true positive (TP), false positive (FP), true negative (TN), and false negative (FN) (24, 25). These assessment measures are used to quantify a predictive model's performance.

For binary classes of datasets A2, B, and C.

$$Acc = \frac{C_P/G_P + C_N/G_N}{G_P + G_N} * 100 \quad (1)$$

$$MCR = \frac{C_N/G_P + C_P/G_N}{N} * 100 \quad (2)$$

$$Sen = \frac{C_P/G_P}{C_P/G_P + C_P/G_N} * 100 \quad (3)$$

$$Spe = \frac{C_N/G_N}{C_N/G_N + C_N/G_P} * 100 \quad (4)$$

$$FPR = \frac{C_N/G_P}{C_N/G_N + C_N/G_P} * 100 \quad (5)$$

$$FNR = \frac{C_P/G_N}{C_P/G_P + C_P/G_N} * 100 \quad (6)$$

The proposed system classifies datasets A, B, and C into two and three classes, namely benign, malignant, and normal. This work trained datasets on multiple epochs like 10, 30, and 50, and the

FIGURE 5 | Image classification of dataset A.

FIGURE 6 | Image classification of dataset B.

Benign
Benign
Benign
Benign
Malignant
Benign
Benign
Benign

TABLE 9 | Confusion matrix of dataset "A" (testing).

Breast cancer	Benign	Malignant	Normal
Benign	86	0	0
Malignant	0	42	0
Normal	1	0	27

TABLE 10 | Confusion matrix of dataset B (testing).

Breast cancer	Benign	Malignant
Benign	476	32
Malignant	20	1,029

TABLE 11 | Confusion matrix of dataset C (testing).

Breast cancer	Benign	Malignant
Benign	108	1
Malignant	1	115

TABLE 12 | Confusion matrix of dataset A2 (testing).

Breast cancer	Benign	Malignant
Benign	87	0
Malignant	0	42

best accuracy of the proposed model is 99.4% for dataset "A," 96.7% for dataset B, 99.1% for dataset C, and 100% for dataset A2 on 50 epochs and 0.0008 learning rate. The model proposed for classification and identification of the BC provided improved accuracy as compared to the earlier work of dataset A (10), their accuracy was 94% of dataset B (11), their accuracy was 80% to 85% and no previous work on dataset C, and the proposed model also achieved 100% on dataset A with two classes (dataset A2).

The algorithm is trained on multiple parameters. Transfer learning-based parameters are utilized for training this model and to get the required output in the proposed system. On 10, 30, and 50 epochs to attain optimal accuracy and loss rate, this study trained the model multiple times. **Table 4** shows the dataset A classes (benign, malignant, and normal accuracy, respectively, 98.9, 100, and 100% and miss rate, respectively, 1.1%, 0.0%, and 0.0%), dataset B classes (benign and malignant accuracy, respectively, 96.0 and 97.0% and miss rate, respectively, 4.0 and 3.0%), dataset C classes (benign and malignant accuracy, respectively, 99.1 and 99.1% and miss rate, respectively, 0.9 and 0.9%), and dataset A2 classes (benign and malignant accuracy, respectively, 100 and 100% and miss rate, respectively, 0.0 and 0.0%) on 50 epochs.

Table 5 shows the accuracy and miss rate of dataset "A" on 10, 30, and 50 epochs. Accuracy is 70.5%, miss rate is 29.5% on 10 epochs, accuracy is 96.8%, miss rate is 3.2% on 30 epochs, and accuracy is 96.8%, miss rate is 3.2% on 50 epochs. **Table 6** shows the miss rate and accuracy of dataset B on 10, 30, and 50 epochs. Accuracy is 77.5%, miss rate is 22.5% on 10 epochs, accuracy is 95.6%, miss rate is 4.4% on 30 epochs, and accuracy is 96.7%, miss rate is 3.3% on 50 epochs. **Table 7** shows the accuracy and miss rate of dataset C on 10, 30, and 50 epochs. Accuracy is 96.0%, miss rate is 4.0% on 10 epochs, accuracy is 97.3%, miss rate is 2.7% on 30 epochs, and accuracy is 99.1%, miss rate is 0.9% on 50 epochs. **Table 8** shows the accuracy and miss rate of dataset A2 on 10, 30, and 50 epochs. Accuracy is 89.1%, miss rate is 10.9% on 10 epochs, accuracy is 96.1%, miss rate is 3.9% on 30 epochs, and accuracy is 100%, miss rate is 0.0% on 50 epochs.

Figure 5 represents the proposed system's labeled pictures of BC according to the dataset "A" classes benign, malignant, and normal. **Figure 6** represents according to the dataset B classes benign and malignant. **Figure 7** represents according to the dataset C classes benign and malignant. **Figure 8** represents according to the dataset A2 classes benign and malignant.

Table 9 shows the confusion matrix of breast cancer classification dataset "A" on 50 epochs. The total number of photographs used for 50 epochs was 780 of dataset "A" with 624 images used for training and 156 images used for testing. A total of 86 images of benign were used for classification in which 86 were classified as benign, 42 images of malignant were used in the classification in which 42 were classified as malignant, and 28 images of normal were used in the classification in which 27 were classified as benign and 1 as benign.

Table 10 shows the confusion matrix of breast cancer classification dataset B on 50 epochs. The total number of photographs used for 50 epochs was 7,783 of dataset A with 6,226 images used for training and 1,557 images used for testing. A total of 508 images of benign were used for classification in which 476 were classified as benign and 32 as malignant, and 1,049 images of malignant were used for classification in which 1,029 were classified as malignant and 20 as benign.

Table 11 shows the confusion matrix of breast cancer classification dataset C on 50 epochs. The total number of photographs used for 50 epochs was 1,126 of dataset C with 225 images used for training and 901 images used for testing. A total of 109 images of benign were used for classification in which 108 were classified as benign and 1 as malignant, and 116 images of malignant were used for classification in which 115 were classified as malignant and 1 as benign.

Table 12 shows the confusion matrix of breast cancer classification dataset A2 on 50 epochs. The total number of photographs used for 50 epochs was 647 of dataset A2 with 518 images used for training and 129 images used for testing. A total of 87 images of benign were used for classification in which 87 were classified as benign and 0 as malignant, and 42 images of malignant were used for classification in which 42 were classified as malignant and 0 as benign.

Figure 9 represents the training accuracy plot which is made up of iterations and epochs and displays the results for 50 epochs of dataset "A." The precision was initially modest, but as the number of epochs increased, it gradually improved. The proposed system is trained at a learning rate of 0.0008 and a total number of six repetitions for each epoch. The chart depicts the

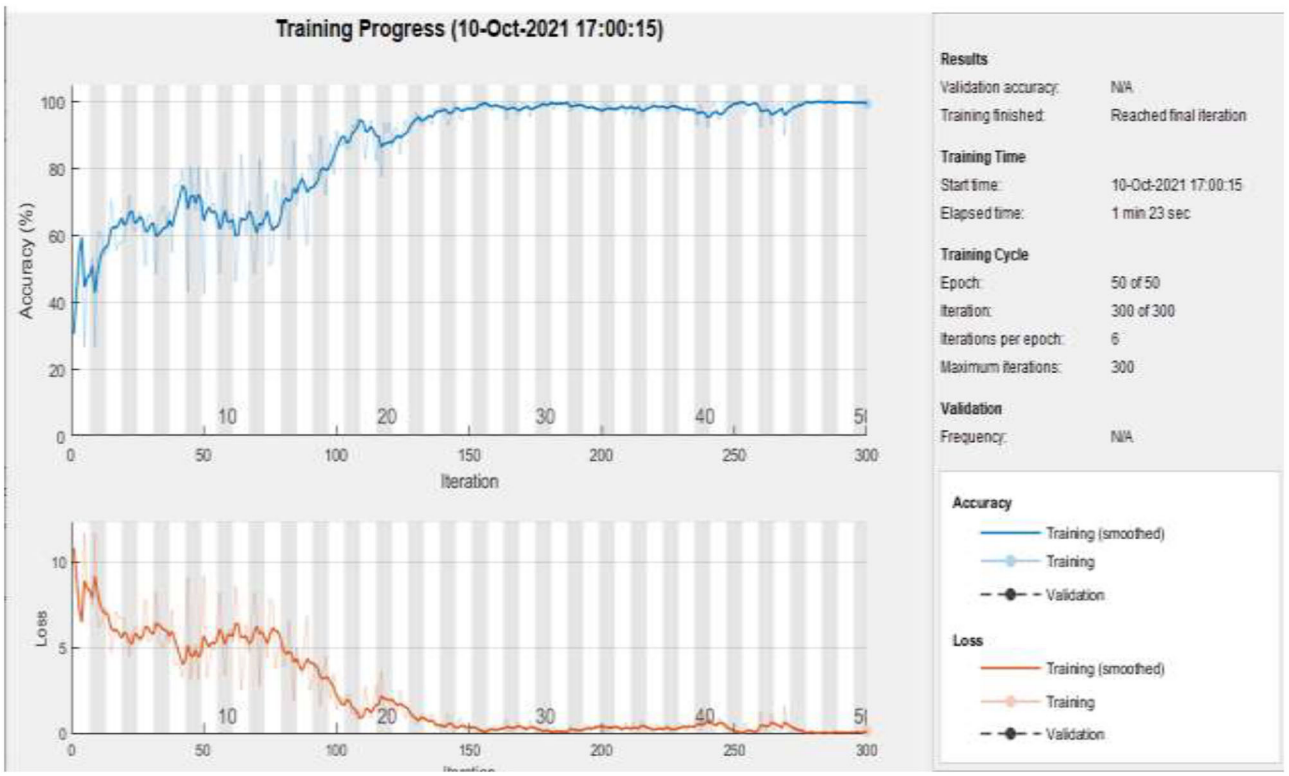


FIGURE 9 | Training of dataset A at 50 epochs.

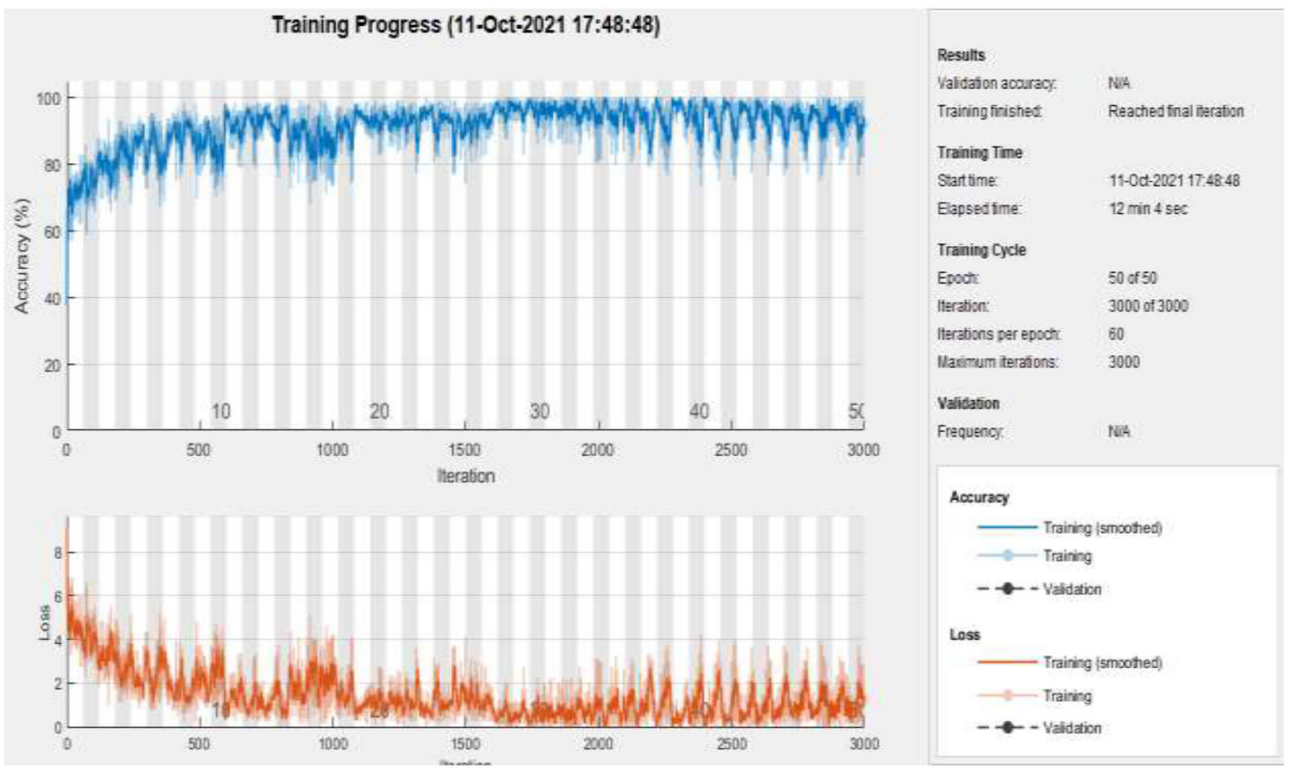


FIGURE 10 | Training of dataset B at 50 epochs.

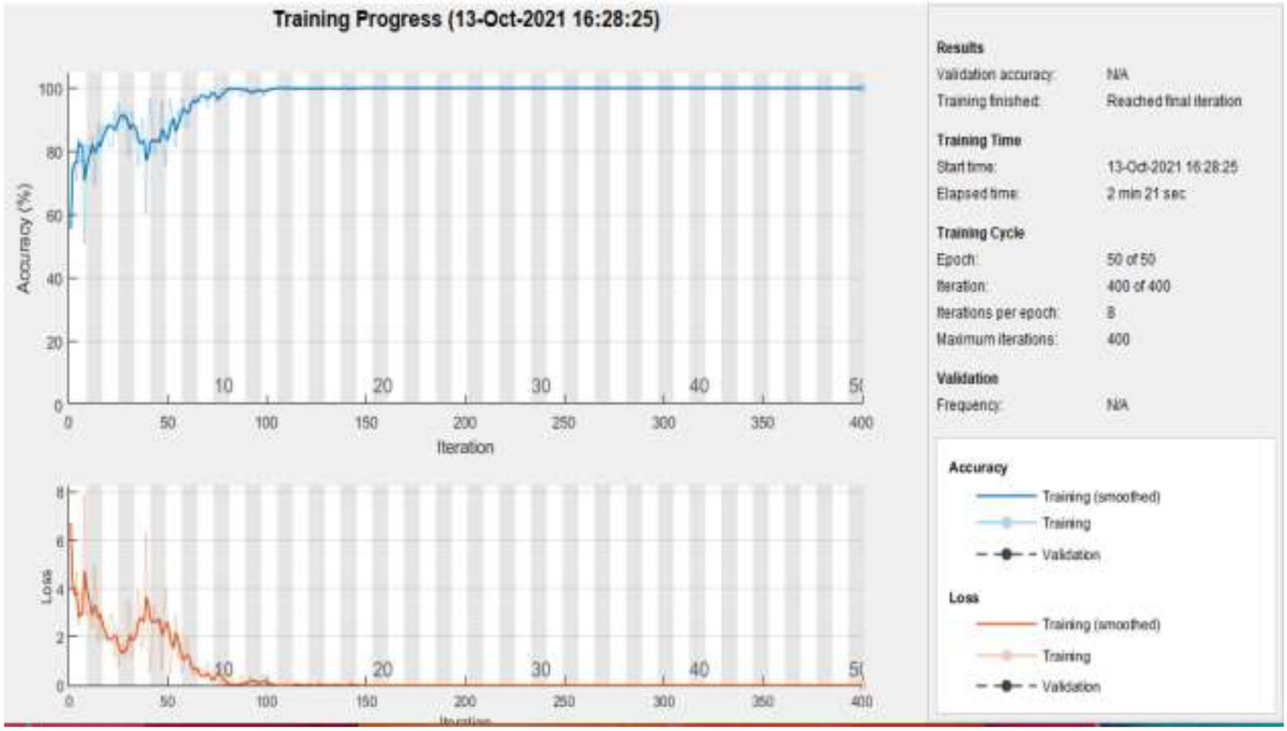


FIGURE 11 | Training of dataset C at 50 epochs.

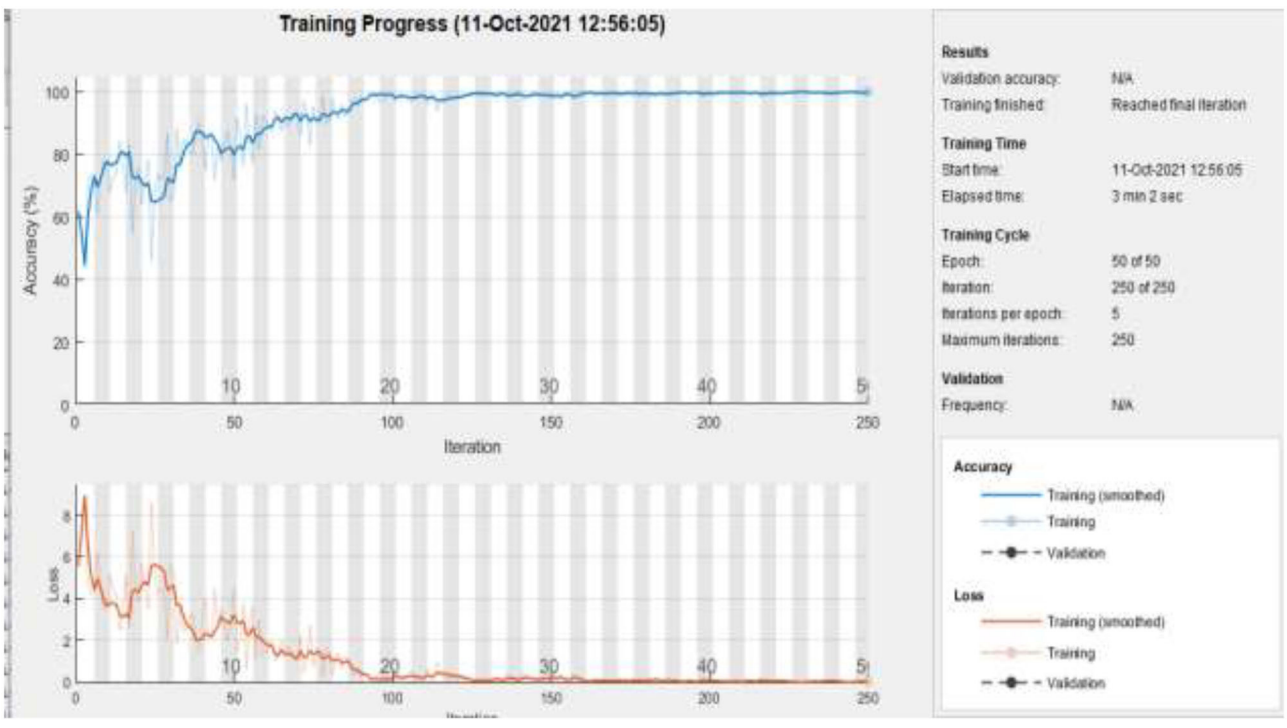


FIGURE 12 | Training of dataset A2 at 50 epochs.

percentage of accuracy for training that began at 1 epoch and ended at 50 epochs.

Figure 10 represents the training accuracy plot which is made up of iterations and epochs and displays the results for 50 epochs of dataset B. The precision was initially modest, but as the number of epochs increased, it gradually improved. The proposed system is trained at a learning rate of 0.0008 and a total number of 60 repetitions for each epoch. The chart depicts the percentage of accuracy for training that began at 1 epoch and ended at 50 epochs.

Figure 11 represents the training accuracy plot which is made up of iterations and epochs and displays the results for 50 epochs of dataset C. The precision was initially modest, but as the number of epochs increased, it gradually improved. The proposed system is trained at a learning rate of 0.0008 and a total number of eight repetitions for each epoch. The chart depicts the percentage of accuracy for training that began at 1 epoch and ended at 50 epochs.

Figure 12 represents the training accuracy plot which is made up of iterations and epochs and displays the results for 50 epochs of dataset A2. The precision was initially modest, but as the number of epochs increased, it gradually improved. The proposed system is trained at a learning rate of 0.0008 and a total number of five repetitions for each epoch. The chart depicts the percentage of accuracy for training that began at 1 epoch and ended at 50 epochs.

The proposed model gives the more precise results as shown in **Table 13**, and it gives TP of class benign 86 from 87, malignant 42 from 42, and normal 26 from 27. The proposed model gave precise results which are shown in **Table 14**; on dataset "A," it gives 99.35% accuracy and 1.149 % FNR for class benign, 100% accuracy and 0.0 % FNR for class malignant, and 99.35% accuracy and 0.0 % FNR for class normal. **Table 15** shows that the proposed model gives 96.66% accuracy and 4.03% FNR on dataset B, 99.11% accuracy and 0.9174%

TABLE 13 | TP, FP, FN, and TN of dataset A.

Data set A	TP	FP	FN	TN
Benign	86	0	1	69
Malignant	42	0	0	115
Normal	27	1	0	128

TABLE 14 | Statistical measures of dataset A.

Classes of dataset A	Acc	MCR	Sen	Spe	FPR	FNR
Benign	99.35%	0.64%	98.85%	100%	0%	1.149%
Malignant	100%	0%	100%	100%	0%	0%
Normal	99.35%	0.64%	100%	99.22%	0.0077%	0%

TABLE 15 | Statistical measures of datasets B, C, and A2.

Data sets	Acc	MCR	Sen	Spe	FPR	FNR
Dataset B	96.66%	3.339%	95.96%	96.98%	3.016%	4.03%
Dataset C	99.11%	0.8888%	99.082%	99.13%	0.8620%	0.9174%
Dataset A2	100%	0%	100%	100%	0%	0%

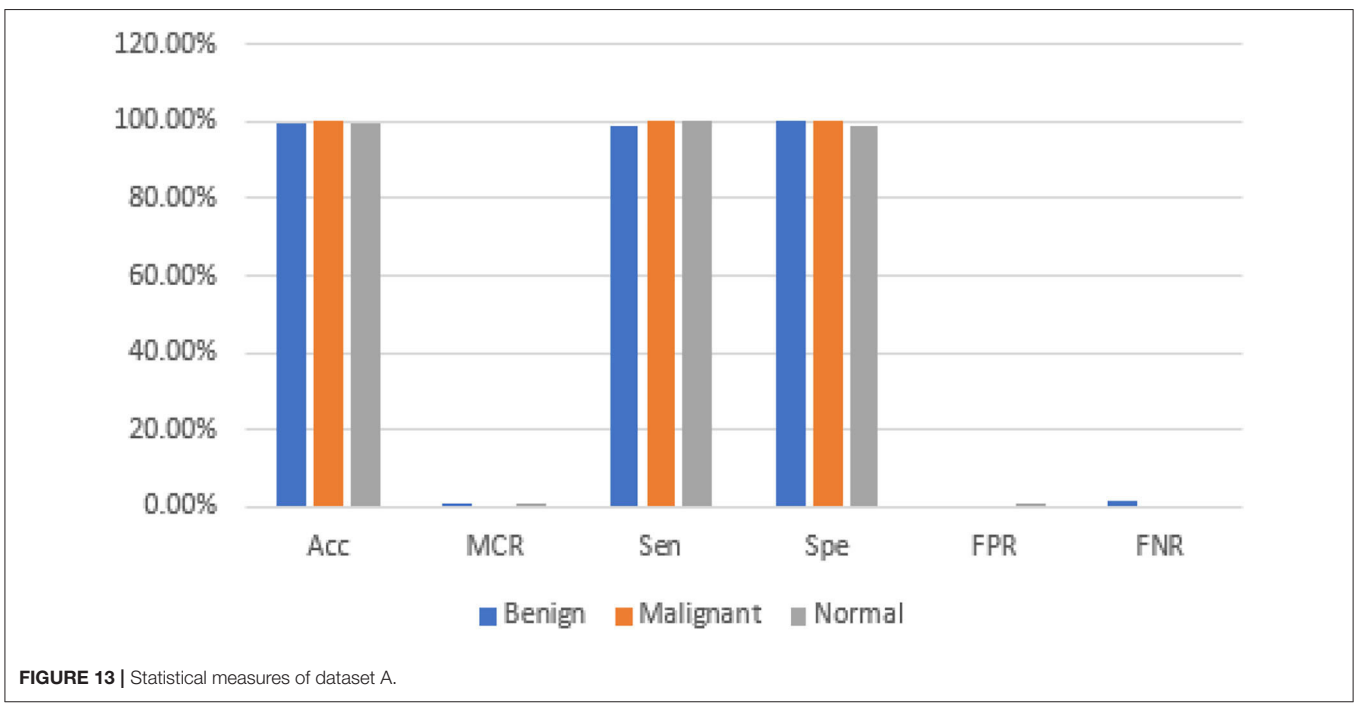


FIGURE 13 | Statistical measures of dataset A.

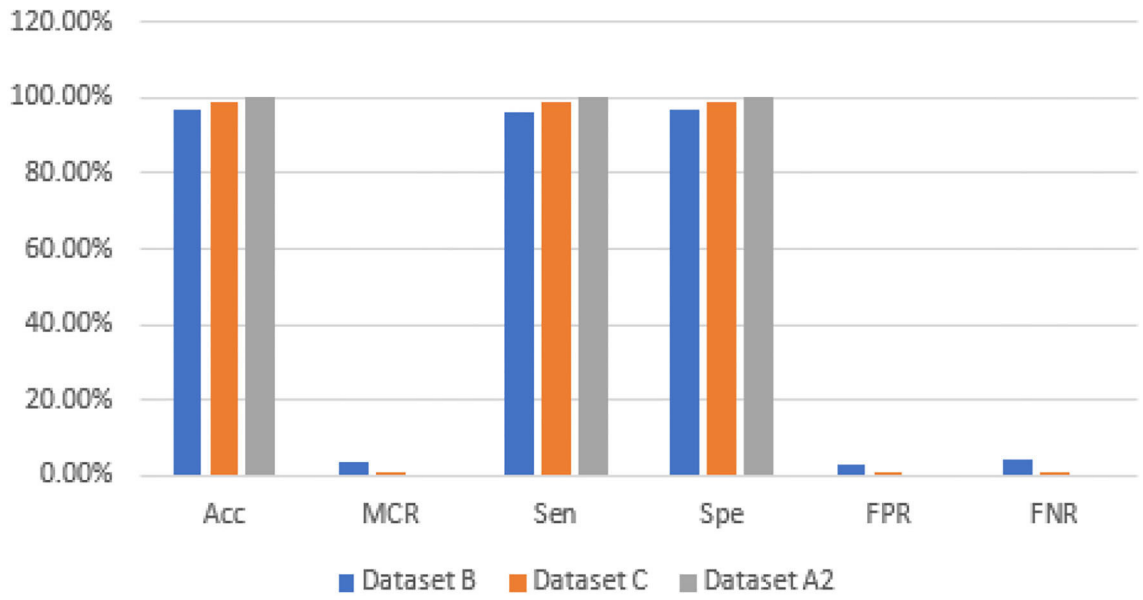


FIGURE 14 | Statistical measures of datasets B, C, and A2.

FIGURE 15 | Accuracy and miss rate in contrast with previous studies.

FNR on dataset C, and 100% accuracy and 0.0% FNR on dataset A2.

The graphical representations of the statistical measures of dataset "A" are shown in **Figure 13**, and the graphical representations of the statistical measures of datasets B, C, and A2 are shown in **Figures 14**, **15**. Multiple methods for detecting BC have been utilized in the past. In the identification of a disease for a given class, the proposed methodology attained good accuracy. As a result, early disease diagnosis can assist medical experts in providing treatment to prevent breast cancer spread. **Table 15** shows the comparison of the suggested system model's performance with the literature work in terms of accuracy and miss rate. The proposed model obtained an accuracy and miss rate of 99.4% and 0.6%, respectively, on dataset "A," accuracy and miss rate of 96.66 and 3.34%, respectively, on dataset B, accuracy and miss rate of 99.11 and 0.89%, respectively, on dataset C, and accuracy and miss rate of 100% and 0%, respectively, on dataset A2. These results show that the proposed model achieved accuracy more than the previous models such as AlexNet, VGG 16, Inception, Res net, and NASNet (10) on dataset BUSI and B, SVM (11), AlexNet (12) on dataset A and B, Inception V3, Res net 50, VGG 16, VGG 19, and Inception V2 Res net (14).

CONCLUSION

The early detection and classification of breast cancer help to prevent the disease's spread. The use of transfer learning AlexNet on breast cancer classification and detection was examined in this work. Deep learning and transfer learning approaches are adapted to the specific properties of any dataset. The proposed model used the customized AlexNet technique on three datasets, A, B, C, and A2, A2 is the dataset A with two classes. This proposed model empowered with transfer learning achieved the best results by using the customized AlexNet. Dataset A has a maximum accuracy of 99.4%, whereas dataset B has a maximum accuracy of 96.70%, dataset C has a maximum accuracy of 99.10%, and dataset A2 has a maximum accuracy of 100%. In future work, we will apply fusion on these datasets for optimum results. We will also apply other CNN algorithms and our model of machine learning on these datasets.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

SA, A-u-R, and MFK have collected data from different resources. SA, MZ, and MAK performed formal analysis and simulation. SA, MFK, KA, and MZ contributed to writing—original draft preparation. A-u-R, MZ, and MAK contributed to writing—review and editing. MAK and AM performed supervision. SA, KA, and MFK drafted pictures and tables. MZ and AM performed revisions and improve the quality of the draft. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. World Health Organization, Cancer. (2022). Available online at: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed September 1, 2021).
2. Saeed S, Jhanjhi NZ, Naqvi M, Humyun M, Ahmad M. Optimized breast cancer pre-mature detection method with computational segmentation: a systematic review mapping. *Approach Appl Deep Learn Virt Med Care*. (2022) 5:24–51. doi: 10.4018/978-1-7998-8929-8.ch002
3. Zhang X, Zhang Y, Zhang Q, Ren Y, Qiu T. Extracting comprehensive clinical information for breast cancer using deep learning methods. *Int J Med Inform*. (2019) 132:103985. doi: 10.1016/j.ijmedinf.2019.103985
4. Swain M, Kisan S, Chatterjee JM, Supramaniam M, Mohanty SN, Jhanjhi NZ, et al. Hybridized machine learning based fractal analysis techniques for breast cancer classification. *Int J Adv Comput Sci Appl*. (2020) 11:179–84. doi: 10.14569/IJACSA.2020.0111024
5. Nanglia S, Ahmad M, Khan FA, Jhanjhi NZ. An enhanced predictive heterogeneous ensemble model for breast cancer prediction. *Biomed Signal Process Control*. (2022) 72:103279. doi: 10.1016/j.bspc.2021.103279
6. Huynh B, Drukker K, Giger M. Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. *Med Phys*. (2016) 43:3705. doi: 10.1118/1.4957255
7. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. (2012) 25:1097–105. doi: 10.1145/3065386
8. Scarpazza C, Baecker L, Vieira S, Mechelli A. Applications of machine learning to brain disorders. *Mach Learn*. (2020) 3:45–65. doi: 10.1016/B978-0-12-815739-8.00003-1
9. Deepak S, Ameer PM. Automated categorization of brain tumor from MRI using CNN features and SVM. *J Ambient Intell Humaniz Comput*. (2021) 12:8357–69. doi: 10.1007/s12652-020-02568-w
10. Dhabyani WA, Gomaa M, Khaled H, Aly F. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *Int J Adv Comput Sci Appl*. (2019) 10:1–11. doi: 10.14569/IJACSA.2019.0100579
11. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng*. (2015) 63:1455–62. doi: 10.1109/TBME.2015.2496264
12. Yap MH, Pons G, Martí J, Ganau S, Sentís M. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform*. (2017) 22:1218–26. doi: 10.1109/JBHI.2017.2731873
13. Ghazal TM, Abbas S, Munir S, Khan MA, Ahmad M. Alzheimer's disease detection empowered with transfer learning. *Comput Mater Contin*. (2021) 70:5005–19. doi: 10.32604/cmc.2022.020866
14. Saber A, Sakr M, Seida OMA, Keshk A, Chen H. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. *IEEE Access*. (2021) 9:71194–209. doi: 10.1109/ACCESS.2021.3079204
15. Azevedo V, Silva C, Dutra I. Quantum transfer learning for breast cancer detection. *Quantum Mach Intell*. (2022) 4:1–14. doi: 10.1007/s42484-022-00062-4
16. Dewangan KK, Dewangan DK, Sahu SP, Janghel R. Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique. *Multimed Tools App*. (2022) 81:13935–60. doi: 10.1007/s11042-022-12385-2
17. Rasool A, Bunterngchit C, Tiejian L, Islam MR, Qu Q, Jiang Q. Improved machine learning-based predictive models for breast cancer Diagnosis. *Int J Environ Res Public Health*. (2022) 19:3211. doi: 10.3390/ijerph19063211
18. Lin RH, Kujabi BK, Chuang CL, Lin CS, Chiu CJ. Application of deep learning to construct breast cancer diagnosis model. *Appl Sci*. (2022) 12:1957. doi: 10.3390/app12041957
19. Alruwaili M, Gouda W. Automated breast cancer detection models based on transfer learning. *Sensors*. (2022) 22:876. doi: 10.3390/s22030876

20. Alshammari MM, Almuhanna A, Alhiyafi J. Mammography image-based diagnosis of breast cancer using machine learning: a pilot study. *Sensors*. (2021) 22:203. doi: 10.3390/s22010203
21. Castro-Tapia S, Castaneda-Miranda CL, Olvera-Olvera CA, Guerrero-Osuna HA, Ortiz-Rodriguez JM, Martinez Blanco MR, et al. Classification of breast cancer in mammograms with deep learning adding a fifth class. *Appl Sci*. (2021) 11:11398. doi: 10.3390/app112311398
22. Mashudi NA, Rossli SA, Ahmad N, Noor NM. Breast cancer classification: features investigation using machine learning approaches. *Int J Integr Eng*. (2021) 13:107–18. doi: 10.30880/ijie.2021.13.05.012
23. Islam MA, Tripura D, Dutta M, Shuvo MNR, Fahim WA, Sarkar PR, et al. Forecast breast cancer cells from microscopic biopsy images using big transfer (BiT): a deep learning approach. *Int J Adv Comput Sci Appl*. (2021) 12:478–86. doi: 10.14569/IJACSA.2021.0121054
24. Ahmed U, Issa GF, Khan MA, Aftab S, Khan MF, Said RAT, et al. Prediction of diabetes empowered with fused machine learning. *IEEE Access*. (2022) 10:8529–38. doi: 10.1109/ACCESS.2022.3142097
25. Siddiqui SY, Haider A, Ghazal TM, Khan MA, Naseer I, Abbas S, et al. IoMT cloud-based intelligent prediction of breast cancer stages empowered with deep learning. *IEEE Access*. (2021) 9:146478–91. doi: 10.1109/ACCESS.2021.3123472
26. Kaggle. (2021) Available online at: https://www.kaggle.com/mostafaeltalawy/brest-cancer (accessed September 1, 2021).
27. Kaggle. (2021) Available online at: https://www.kaggle.com/anaselmasry/breast-cancer-dataset (accessed September 1, 2021).
28. Kaggle. (2021) Available online at: https://www.kaggle.com/aryashah2k/breast-ultrasound-images-dataset (accessed September 1, 2021).
29. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
30. Arooj S, Khan MF, Khan MA, Khan MS, Taleb N. Machine Learning Models for the Classification of Skin Cancer. In: *International Conference on Business Analytics for Technology and Security (ICBATS)* (IEEE) (2022). p. 1–8. doi: 10.1109/ICBATS54253.2022.9759054
31. Alom MZ. The history began from alexNet: a comprehensive survey on deep learning approaches. *Arxiv*. (2018) 18:1–12. doi: 10.48550/arXiv.1803.01164

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Arooj, Atta-ur-Rahman, Zubair, Khan, Alissa, Khan and Mosavi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 06 July 2022
doi: 10.3389/fpubh.2022.922510

Using Machine Learning Methods to Predict Bone Metastases in Breast Infiltrating Ductal Carcinoma Patients

Wen-Cai Liu [1,2†], Ming-Xuan Li [2†], Shi-Nan Wu [2†], Wei-Lai Tong [1,3], An-An Li [1,3], Bo-Lin Sun [1,3], Zhi-Li Liu [1,3] and Jia-Ming Liu [1,3]*

[1] Department of Orthopaedic Surgery, The First Affiliated Hospital of Nanchang University, Nanchang, China, [2] Department of Clinical Medicine, The First Clinical Medical College of Nanchang University, Nanchang, China, [3] Institute of Spine and Spinal Cord, Nanchang University, Nanchang, China

Breast cancer (BC) was the most common malignant tumor in women, and breast infiltrating ductal carcinoma (IDC) accounted for about 80% of all BC cases. BC patients who had bone metastases (BM) were more likely to have poor prognosis and bad quality of life, and earlier attention to patients at a high risk of BM was important. This study aimed to develop a predictive model based on machine learning to predict risk of BM in patients with IDC. Six different machine learning algorithms, including Logistic regression (LR), Naive Bayes classifiers (NBC), Decision tree (DT), Random Forest (RF), Gradient Boosting Machine (GBM), and Extreme gradient boosting (XGB), were used to build prediction models. The XGB model offered the best predictive performance among these 6 models in internal and external validation sets (AUC: 0.888, accuracy: 0.803, sensitivity: 0.801, and specificity: 0.837). Finally, an XGB model-based web predictor was developed to predict risk of BM in IDC patients, which may help physicians make personalized clinical decisions and treatment plans for IDC patients.

Keywords: breast cancer, infiltrating ductal carcinoma, bone metastases, machine learning, prediction

OPEN ACCESS

Edited by:
Ali Kashif Bashir, Manchester Metropolitan University, United Kingdom

Reviewed by:
Chuan-Gui Song, Affiliated Union Hospital of Fujian Medical University, China
Sina Ardabili, University of Mohaghegh Ardabili, Iran

***Correspondence:**
Jia-Ming Liu
liujiamingdr@hotmail.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *18 April 2022*
Accepted: *16 June 2022*
Published: *06 July 2022*

Citation:
Liu W-C, Li M-X, Wu S-N, Tong W-L, Li A-A, Sun B-L, Liu Z-L and Liu J-M (2022) Using Machine Learning Methods to Predict Bone Metastases in Breast Infiltrating Ductal Carcinoma Patients. Front. Public Health 10:922510. doi: 10.3389/fpubh.2022.922510

INTRODUCTION

As one of the most common malignancies, breast cancer (BC) accounted for 30% of all cancers in women (1). The incidence of BC continued to increase at a rate of ∼0.5% per year, which was attributed at least in part to the continued decline in fertility and increased body weight (2). In the cancer statistical report, the number of BC patients exceeded 2.1 million, and infiltrating ductal carcinoma (IDC) was the most common one among different types of BC (3, 4). Early diagnosis provides a favorable prognosis and better overall survival for BC patients. In North America, early screening for BC significantly increased the 5-year survival rate to over 80% for BC patients (5). In recent years, BC patients still had a high incidence of distant metastatic recurrence, which was an important indicator for poor prognosis. Recent studies have suggested that age, menopausal status, T, N stages, histological grade and HR/HER2 status were risk factors for BM in BC patients (6–8). For distant metastases, bone was the most common site, and more than 60% of BC patients developed bone metastases (BM) (9). Another study indicated over 20% of patients developed BM within an average follow-up duration of 8 years. And further survival prognostic modeling showed a 40-month median survival time for patients with BM (10). In addition, BC patients with BM

often developed secondary clinical complexities that took up significant medical resources (11). Early diagnosis of BM from BC will help in the timely prevention and treatment of complications and increase the quality of life for BC patients.

Currently, precision medicine has preceded four concepts: predictive, personalized, preventive and participatory (12). The technology of big data analytics is becoming a clinical imperative (13). This means that we need to use advanced technology to analyze large amounts of medical data to provide recommendations for individualized treatment. Many studies have used machine learning (ML) techniques to study clinical risk factors associated with cancer metastases for early detection (14–16). It is known that the most common type of pathology in BC is IDC (4). But few studies focused on incorporating machine learning to predict the risk of BM from IDC patients.

In this study, we attempted to develop a predictive model to predict the BM risk in IDC patients based on machine learning, and to assist clinicians in implementing more rational clinical decisions as well as to enable patients to receive earlier treatment. Our contribution includes:

- Machine learning algorithms were used instead of traditional statistical regression methods to process data on clinical characteristics of IDC patients to identify patients at high risk for BM.
- This paper compared various algorithms that could be used to process patient data and identified XGB as the best method for processing the data. Further, the hyperparameters of the XGB algorithm were fine-tuned using a random search method to improve performance.
- We performed importance analysis of the features included in the model using the permutation importance method, and these features were further analyzed and understood from a clinical medicine perspective in the discussion section.
- This study proposed a machine learning based solution that could assist clinicians in making individualized diagnoses of BM for IDC patients.

The rest of the paper was organized as follows: the materials and methods section provided a detailed documentation of the materials and methods used, as well as the dataset description, statistical analyses, data pre-processing, feature engineering, classification algorithms used and evaluation metrics. Results of the experiment are discussed in Section Results and then further discussed in Section Discussion. Conclusions, limitations and future work section concluded the results and provided limitations and the future direction of the current work.

MATERIALS AND METHODS

Study Population

The training and internal test set data was derived from the Surveillance, Epidemiology, and End Results (SEER) database, and the external test set data was derived from the First Affiliated Hospital of Nanchang University in China. A data set of 311,408 IDC patients was included from the SEER database (2010–2017) and sliced into a training set and an internal test set randomly in a ratio of 7:3. An external validation set included data from 1,243 IDC patients of our hospital (2010–2017). The exclusion criteria for clinical data were as follows: (1) unknown information of T, N stage, race, laterality, breast subtype, grade and marital status. (2) Other cases with unknown primary tumor and metastatic status. The detailed exclusion criteria were shown in **Figure 1**. Information of all variables was complete for these patients.

Data Selection

We selected nine variables from the SEER database and our hospital that may affect BM in patients with IDC, including age at diagnosis, race, sex, grade, T, N stage, breast subtype, laterality and marital status. All cases included in this study were staged using the 7th edition of the AJCC TNM staging system and the relevant guidelines of the SEER project.

Statistical Analyses

The statistical analyses in this study were all performed by Python (version 3.8, Python Software Foundation) and SPSS (version 26, IBM, USA). Data from SEER were randomly sliced into training and internal test sets in a ratio of 7:3 using python. The training set was used to construct the models, and the internal test set and external test set were used for model validation and evaluation. A heat map was drawn to determine the association among the variables. A univariate analysis was performed to compare variables between patients with and without BM. For categorical data, the chi-square test was used, and for continuous non-normally distributed data, the Wilcoxon rank-sum test was used. Variables with a $P < 0.05$ in univariate analysis were enclosed within the construction of machine learning models and multivariate logistic regression was performed to identify the risk factors of BM from IDC patients.

Data Pre-processing and Feature Engineering

Category variables such as T and N stages were processed using label encoding methods. The univariate analysis was used to screen for meaningful combinations of features for predicting the risk of IDC patients with BM. Correlation analysis was used to analyze the correlation among the selected features. Feature importance analysis was performed on the variables based on the Permutation Importance principle (17, 18).

Evaluation Metrics

The purpose of this study was to accurately predict the clinical outcome of a specific patient based on multiple variables. So the predictive power and accuracy of the model were important. Thus, to evaluate the model, we considered the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity (recall rate) and specificity score in the study. The following terms were used in the equations: TP (True Positive); TN (True Negative); FP (False Positive); and FN (False Negative).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$sensitivity\ (recall\ rate) = \frac{TP}{TP + FN}$$

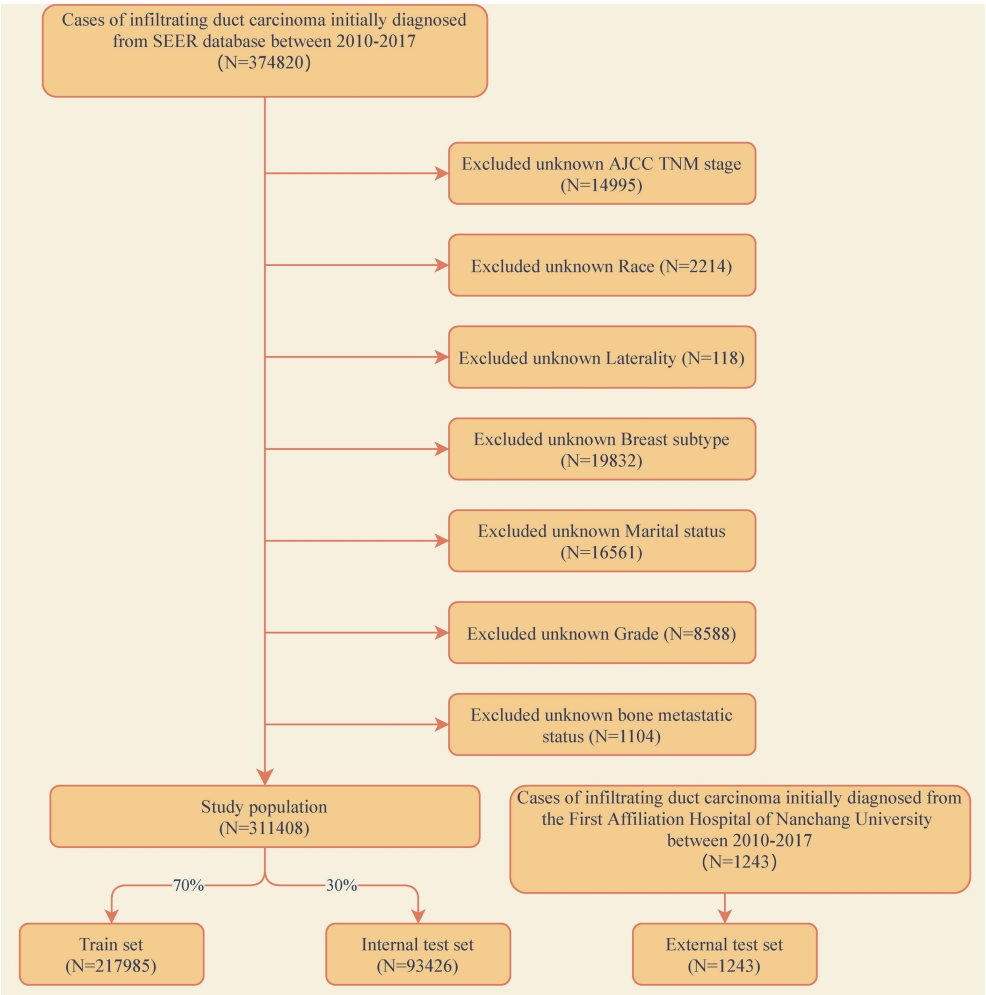


FIGURE 1 | Flow diagram of the study population selected from the Surveillance, Epidemiology, and End Results (SEER) database and the First Affiliated Hospital of Nanchang University. According to the inclusion and exclusion criteria, a total of 311,408 patients of SEER were included in this study, and they were randomly cut into the training and internal test sets in a 7:3 ratio. Data from the First Affiliated Hospital of Nanchang University (n = 1,243) as an external test set.

$$specificity = \frac{TN}{TN + FP}$$

Model Establishment

All the algorithmic models were built based on scikit-learn (version 0.24.2). The random oversampling method in imbalanced-learn (version 0.9.1) was used to deal with the imbalance of data distribution.

In this study, we used six different machine learning algorithms: Logistic regression (LR), Naive Bayes classifiers (NBC), Decision tree (DT), Random Forest (RF), Gradient Boosting Machine (GBM) and Extreme gradient boosting (XGB) (19–24). The ML algorithms were trained and adjusted to predict the BM in IDC patients. Random search method in scikit-learn was used to adjust the hyperparameters of the model. Then, the predictive performance of the ML models was evaluated in internal 10-fold cross-validation of the train set, internal and external test sets and the AUC, accuracy, sensitivity (recall rate) and specificity score were evaluated. Then, we selected the best-performing model to build a web predictor.

RESULTS

Demographic Baseline Characteristics

A total of 311,408 cases who were first diagnosed with IDC in the SEER database from 2010 to 2017 were included. Of these cases, 7,949 (2.55%) complicated with BM and 303,459 (97.45%) without BM. All demographic and clinicopathological characteristics of these patients were demonstrated in detail in **Table 1**. All cases were randomly divided into a training set (n = 217,985) and an internal test set (n = 93,426) in a ratio of 7:3. Data for the external test set were derived from IDC patients firstly diagnosed in our hospital from 2010 to 2017 (N = 1,243). The details of the training and test sets were shown in **Table 2**.

TABLE 1 | Clinical and pathological characteristics of study population.

Variables	ALL	NBM	BM
	N = 311,408	*N* = 303,634	*N* = 7,774
Age			
<50	65,967 (21.2%)	64,129 (21.1%)	1,838 (23.6%)
≥50	245,441 (78.8%)	239,505 (78.9%)	5,936 (76.4%)
Sex			
Female	308,805 (99.2%)	301,162 (99.2%)	7,643 (98.3%)
Male	2,603 (0.8%)	2,472 (0.8%)	131 (1.7%)
Race			
American Indian/Alaska Native	1,848 (0.6%)	1,809 (0.6%)	39 (0.5%)
Asian or Pacific Islander	28,929 (9.3%)	28,312 (9.3%)	617 (7.9%)
Black	35,011 (11.2%)	33,749 (11.1%)	1,262 (16.2%)
White	245,620 (78.9%)	239,764 (79.0%)	5,856 (75.3%)
Grade			
Grade I (well differentiated)	65,791 (21.1%)	65,252 (21.5%)	539 (6.9%)
Grade II (moderately differentiated)	132,463 (42.5%)	128,913 (42.5%)	3,550 (45.7%)
Grade III (poorly differentiated)	112,515 (36.1%)	108,858 (35.9%)	3,657 (47.0%)
Grade IV (undifferentiated)	639 (0.2%)	611 (0.2%)	28 (0.4%)
Breast subtype			
HR-/HER2- (triple negative)	38,740 (12.4%)	37,927 (12.5%)	813 (10.5%)
HR-/HER2+ (HER2 enriched)	15,803 (5.1%)	15,246 (5.0%)	557 (7.2%)
HR+/HER2- (Luminal A)	219,700 (70.6%)	214,764 (70.7%)	4,936 (63.5%)
HR+/HER2+ (Luminal B)	37,165 (11.9%)	35,697 (11.8%)	1,468 (18.9%)
T stage			
T1	191,204 (61.4%)	190,153 (62.6%)	1,051 (13.5%)
T2	93,067 (29.9%)	90,289 (29.7%)	2,778 (35.7%)
T3	15,307 (4.9%)	14,031 (4.6%)	1,276 (16.4%)
T4	11,830 (3.8%)	9,161 (3.0%)	2,669 (34.3%)
N stage			
N0	215,120 (69.1%)	213,308 (70.3%)	1,812 (23.3%)
N1	72,080 (23.1%)	68,326 (22.5%)	3,754 (48.3%)
N2	15,459 (5.0%)	14,400 (4.7%)	1,059 (13.6%)
N3	8,749 (2.8%)	7,600 (2.5%)	1,149 (14.8%)
Laterality			
Left	157,489 (50.6%)	153,472 (50.5%)	4,017 (51.7%)
Right	153,919 (49.4%)	150,162 (49.5%)	3,757 (48.3%)
Marital status			
Married	261,707 (84.0%)	255,829 (84.3%)	5,878 (75.6%)
Unmarried	49,701 (16.0%)	47,805 (15.7%)	1,896 (24.4%)

BM, Bone metastasis; NBM, No bone metastasis; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

Univariate Analysis and Multivariate Logistic Regression Analysis

According to the univariate analysis, patients' age, gender, race, grade, T, N stage, breast subtype and marital status were significantly associated with BM in patients with IDC ($P < 0.05$; **Table 3**). Variables with a $P < 0.05$ in the univariate analysis were selected for multivariate logistic regression analysis, in order to identify the risk factors of BM in IDC patients. According to these results, age, race, grade, T, N stage, breast subtype, and marital status were found to be independent factors for BM (**Table 3**).

Correlation Analysis of Features

To identify the effect of these features on prediction, correlation tests between each other were performed. Correlation analysis among dataset features provides information about the degree of interaction among features. The heat map showed the relevance of these features on the prediction of ML algorithm (**Figure 2**). The figure showed a positive correlation among T-stage, N-stage and pathological grade. This was consistent with clinical experience that poorly differentiated tumor tissue was poorly demarcated from surrounding normal tissue, which meant it was more aggressive.

TABLE 2 | Clinical and pathological characteristics of training set and test set.

Variables	Training set		Internal test set		External test set	
	NBM (%) (n = 212,567)	BM (%) (n = 5,418)	NBM (%) (n = 91,067)	BM (%) (n = 2,356)	NBM (%) (n =1,068)	BM (%) (n = 175)
Age						
<50	45,017 (21.2)	1,270 (23.2)	19,112 (21.0)	568 (24.1)	358 (33.5)	69 (39.4)
≥50	167,550 (78.8)	4,148 (76.8)	71,955 (79.0)	1,788 (75.9)	710 (66.5)	106 (60.6)
Sex						
Female	210,814 (99.2)	5,322 (98.2)	90,348 (99.2)	2,321 (98.5)	1,064 (99.6)	173 (98.9)
Male	1,753 (0.8)	96 (1.8)	719 (0.8)	35 (1.5)	4 (0.4)	2 (1.1)
Race						
American Indian/Alaska Native	1,267 (0.6)	24 (0.4)	542 (0.6)	15 (0.6)	0 (0.0)	0 (0.0)
Asian or Pacific Islander	19,981 (9.4)	415 (7.7)	8,331 (9.1)	202 (8.6)	1,068 (100.0)	175 (100.0)
Black	23,645 (11.1)	888 (16.4)	10,104 (11.1)	374 (15.9)	0 (0.0)	0 (0.0)
White	167,674 (78.9)	4,091 (75.5)	72,090 (79.2)	1,765 (74.9)	0 (0.0)	0 (0.0)
Grade						
Grade I (well differentiated)	45,558 (21.4)	385 (7.1)	19,694 (21.6)	154 (6.5)	202 (18.9)	11 (6.3)
Grade II (moderately differentiated)	90,360 (42.5)	2,461 (45.4)	38,553 (32,631)	1,090 (46.3)	494 (46.3)	82 (46.9)
Grade III (poorly differentiated)	76,227 (35.9)	2,555 (47.2)	32,631 (35.8)	1,102 (46.8)	369 (34.6)	81 (46.3)
Grade IV (undifferentiated)	422 (0.2)	18 (0.3)	189 (0.2)	10 (0.4)	3 (0.3)	1 (0.6)
Breast subtype						
HR-/HER2- (triple negative)	26,586 (12.5)	569 (10.5)	11,341 (12.5)	244 (10.4)	95 (8.9)	9 (5.1)
HR-/HER2+ (HER2 enriched)	10,737 (5.1)	377 (7.0)	4,509 (5.0)	180 (7.6)	73 (6.8)	16 (9.1)
HR+/HER2- (Luminal A)	150,391 (70.7)	3,453 (63.7)	64,373 (70.7)	1,483 (62.9)	765 (71.6)	118 (67.4)
HR+/HER2+ (Luminal B)	24,853 (11.7)	1,019 (18.8)	10,844 (11.9)	449 (19.1)	135 (12.6)	32 (18.3)
T stage						
T1	132,936 (62.5)	750 (13.8)	57,217 (62.8)	301 (12.8)	639 (59.8)	12 (6.9)
T2	63,445 (29.8)	1,929 (35.6)	26,844 (29.5)	849 (36.0)	350 (32.8)	62 (35.4)
T3	9,829 (4.6)	866 (16.0)	4,202 (4.6)	410 (17.4)	52 (4.9)	33 (18.9)
T4	6,357 (3.0)	1,873 (34.6)	2,804 (3.1)	796 (33.8)	27 (2.5)	68 (38.9)
N stage						
N0	149,389 (70.3)	1,261 (23.3)	63,919 (70.2)	551 (23.4)	746 (69.9	41 (23.4)
N1	47,801 (22.5)	2,630 (48.5)	20,525 (22.5)	1,124 (47.7)	241 (22.6)	79 (45.1)
N2	10,019 (4.7)	722 (13.3)	4,381 (4.8)	337 (14.3)	60 (5.6)	30 (17.1)
N3	5,358 (2.5)	805 (14.9)	2,242 (2.5)	344 (14.6)	21 (2.0)	25 (14.3)
Laterality						
Left	107,409 (50.5)	2,763 (51.0)	46,063 (50.6)	1,254 (53.2)	523 (49.0)	95 (54.3)
Right	105,158 (49.5)	2,655 (49.0)	45,004 (49.4)	1,102 (46.8)	545 (51.0)	80 (45.7)
Marital status						
Married	179,137 (84.3)	4,079 (75.3)	76,692 (84.2)	1,799 (76.4)	915 (85.7)	133 (76.0)
Unmarried	33,430 (15.7)	1,339 (24.7)	14,375 (15.8)	557 (23.6)	153 (14.3)	42 (24.0)

BM, Bone metastasis; NBM, No bone metastasis; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.

Relative Feature Importance on Prediction

The feature importance of each ML model for predicting BM was illustrated in **Figure 3**. Permutation importance principle was used to analyze the relative feature importance on the variables in each ML model. The basic idea of the principle was: (1) Train the model. (2) Disrupt the data in one of the columns and use that dataset for prediction, evaluating the decrease in prediction accuracy to reflect the importance of that feature variable. (3) Restore the validation dataset and repeat the second step to analyze the other feature variables. Although the relative feature importance in different ML models varied slightly, T, N stage, breast subtype, grade, and marital status were the top-ranked variables in most models. In contrast, the race was the last one in most models. But it also contributed to the model. In the XGB model, the relative feature importance was sorted in descending order by T, N stage, breast subtype, grade, marital status, laterality, age, sex, and race.

Model Performance

The prediction performance of all models was compared with the internal 10-fold cross-validation of the training set, internal and external test sets, as shown in **Figures 4–6** and **Table 4**. In the

TABLE 3 | Univariate analysis and multivariate logistic regression analysis of variables.

Variables	Univariate analysis		Multivariate logistic regression analysis	
	Chi-Square	***P*-value**	**OR (95%CI)**	***P*-value**
Age	12.612	<0.001		
<50			Reference	
≥50			1.122 (1.053–1.196)	<0.001*
Sex	56.360	<0.001		
Female			Reference	
Male			1.185 (0.946–1.485)	0.140
Race	157.012	<0.001		
American Indian/Alaska Native			Reference	
Asian or Pacific Islander			1.179 (0.764–1.818)	0.456
Black			1.641 (1.071–2.516)	0.023*
White			1.550 (1.016–2.365)	0.042*
Grade	716.162	<0.001		
Grade I (well differentiated)			Reference	
Grade II (moderately differentiated)			1.546 (1.380–1.732)	<0.001*
Grade III (poorly differentiated)			1.201 (1.066–1.353)	0.003*
Grade IV (undifferentiated)			1.314 (0.786–2.197)	0.298
Breast_subtype	317.014	<0.001		
HR-/HER2- (triple negative)			Reference	
HR-/HER2+ (HER2 enriched)			1.314 (1.143–1.510)	<0.001*
HR+/HER2- (Luminal A)			1.767 (1.600–1.951)	<0.001*
HR+/HER2+ (Luminal B)			2.003 (1.793–2.239)	<0.001*
T stage	17,446.197	<0.001		
T1			Reference	
T2			3.379 (3.464–4.149)	<0.001*
T3			8.810 (7.896–9.830)	<0.001*
T4			27.233 (24.635–30.104)	<0.001*
N stage	6,878.970	<0.001		
N0			Reference	
N1			3.061 (2.843–3.295)	<0.001*
N2			2.645 (2.387–2.932)	<0.001*
N3			4.390 (3.953–4.875)	<0.001*
Laterality	0.461	0.497		
Left			-	-
Right			-	-
Marital status	318.307	<0.001		
Married			Reference	
Unmarried			1.363 (1.272–1.460)	<0.001*

BM, Bone metastasis; NBM, No bone metastasis; HR, hormone receptor; HER2, human epidermal growth factor receptor 2.
$^{*}P < 0.05$.

internal test set, the XGB model was the best performer with an AUC of 0.857, an accuracy of 0.787, a sensitivity of 0.787 and a specificity of 0.791. In the external test set, the XGB model showed a relatively better performance with an AUC of 0.888, an accuracy of 0.803, a sensitivity of 0.801 and a specificity of 0.837 (**Table 4**). Finally, we used the highest-performing model to create a web predictor.

Web Predictor

Based on the XGB model, a ML algorithm for optimal predictive performance, the web predictor mentioned above was developed to predict the risk of BM in IDC patients. We can predict the BM risk in IDC patients simply by setting variables in the sidebar of the website (**Figure 7**) (https://share.streamlit.io/liuwencaincu/breast-cancer/main/breast.py).

DISCUSSION

Breast cancer was the most common malignant tumor in women, and IDC accounted for about 80% of all BC cases (3). BC patients who had BM were more likely to have poor prognosis and bad quality of life (25), and earlier attention to patients at a

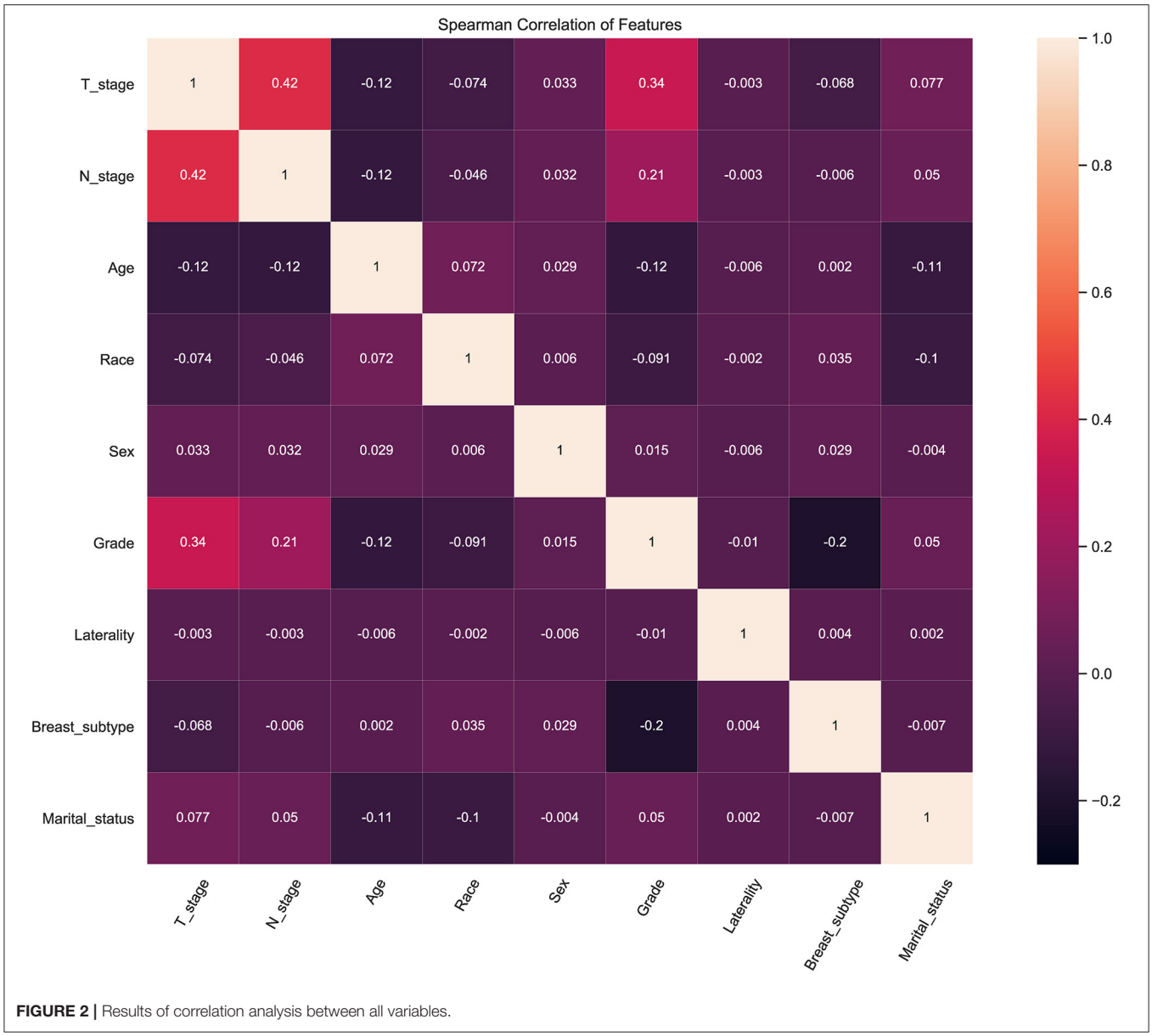


FIGURE 2 | Results of correlation analysis between all variables.

high risk of BM was important. However, metastatic BC was a highly heterogeneous disease, and bone was the most common site of distance metastasis from BC (3, 5, 26). A decline in quality of life was common among patients with BC who suffered from BM. And the risk of skeletal-related events significantly increased following BM in BC patients, such as pain, fracture and hypercalcemia (27, 28). The life expectancy of patients with BC benefitted from early detection and systemic therapy. Therefore, the prognosis of patients with late detection was often poor against a background of increasing incidence of BM.

Bone scintigraphy is often used to distinguish BC patients likely to develop BM. However, this method may not be suitable for early screening due to the expense and radiation damage associated with bone imaging. To help address these potential problems, we built a predictive model using ML technologies to predict BM in IDC patients and identify patients at a high risk of BM.

Presently, new techniques in ML and artificial intelligence (AI) help us succeed in translational research in many fields (29–31). Given the success of ML techniques in other fields, coupled with the large amount of data available in healthcare, ML techniques will have a promising future in the medical field (32–34). In particular, the emergence of electronic medical records (EMR) has generated a large accumulation of clinical data sets, including clinical diagnoses and laboratory data. ML in the medical field can lead to accurate diagnosis and personalized patient care (35). Since the outbreak of the COVID-19, scientists have used ML techniques to develop a variety of predictive and diagnostic models based on big clinical data from patients to help health care professionals work together to address the pandemic

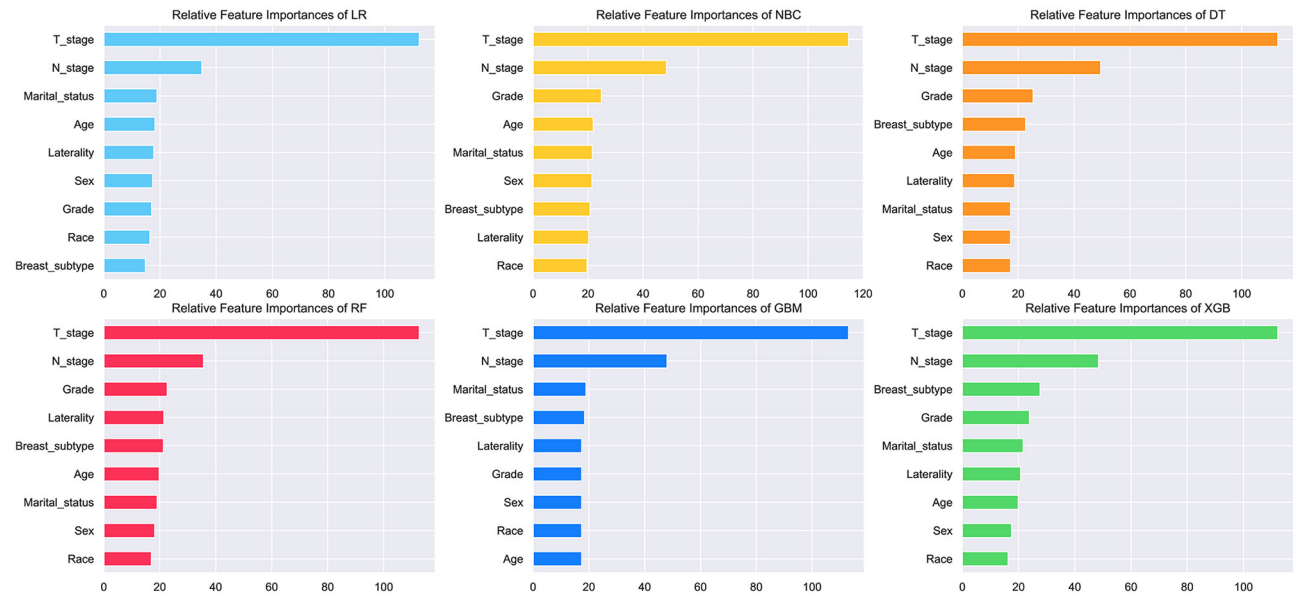


FIGURE 3 | Relative feature importance of different models. The plot showed the ranking of the relative importance of features in all models.

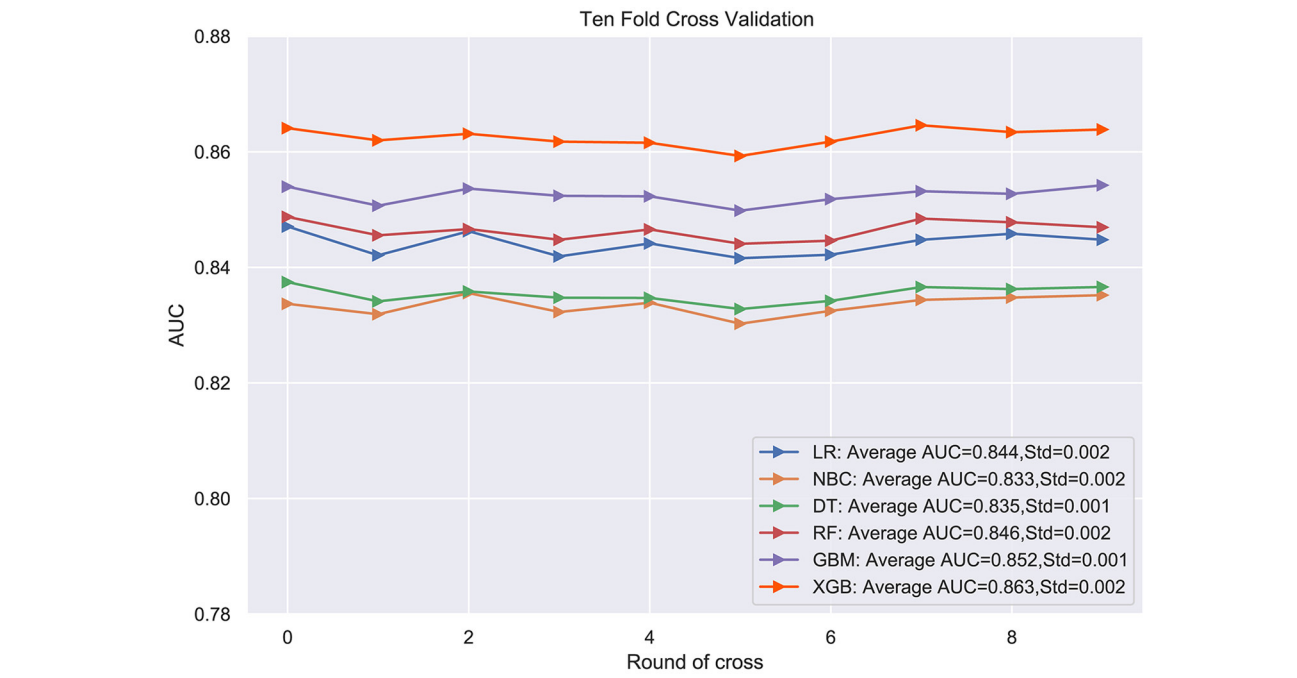


FIGURE 4 | Ten-fold cross-validation results of different machine learning models in the training set. DT, Decision tree; LR, Logistic regression; GBM, Gradient Boosting Machine; NBC, Naive Bayes classification; RF, Random Forest; XGB, Extreme gradient boosting.

(36–38). Statistical and comprehensive reviews of ML in medical diagnosis by Bhavsar et al. (35, 39) suggested that ML techniques can help medical professionals reduce diagnostic errors, improve healthcare services, and cut treatment costs. And in the cancer metastases field, some clinical prediction models in predicting the risk of BM based on ML algorithms have been developed to assist clinicians in personalizing patient diagnosis (40, 41).

This study was novel in using ML algorithms to predict the risk of BM in IDC patients. To our knowledge, researchers in the medical field have only used traditional linear statistical models to predict the risk of BM in IDC patients, and few studies applied ML techniques to them (42). A ML model based on XGB algorithm was developed to accurately predict the risk of BM, outperforming other models developed in this study. The model

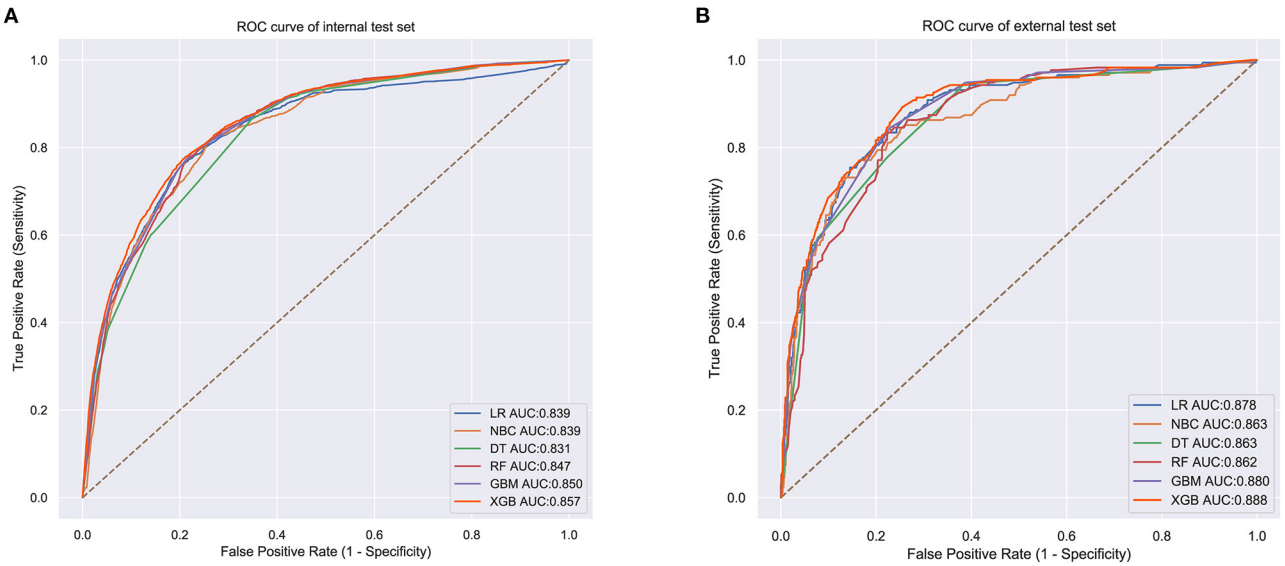


FIGURE 5 | (A) Internal test ROC curve of different machine learning models. **(B)** External test ROC curve of different machine learning models.

established in this study had great discrimination and showed satisfactory specificity and sensitivity.

In this study, several models were constructed and validated to predict BM risk in patients with IDC using current ML methods, and logistic regression analysis demonstrated that age, race, grade, T, N stage, breast subtype, and marital status were independent risk factors for BM. After analyzing the performance of six ML algorithms, we found that the XGB method performed the best (AUC-0.888). The XGB algorithm added a regular term to the objective function to control the complexity of the model and avoid overfitting, while supporting column sampling to enhance the stability of the model (19). This may be part of the reason why it achieved the best performance in our study. To increase the practicality of this model implementation, we created an online web calculator for calculating individual BM probability in IDC patients.

Previous study indicated that age and race were investigated broadly as risk factors for BC metastasis (43). It was reported that white women were less likely to have BM from BC than black women (44). In our study, blacks also had a higher incidence of BM than those of Whites, American Indians and Asia-Pacific people. Chen et al. and Wang et al. found that elderly women were more likely to develop BM from BC (45, 46). In the present study, we found that advanced age was a risk factor for BM in patients with IDC.

Tumor size was positively correlated with the chance of BM. Yazdani et al. found that tumor size was a risk factor of BM from BC and a larger tumor size increased the likelihood of BM (47). In this study, patients with N3 stage were significantly more likely to develop BM than patients with other stages. And the average N stage in patients with BM was higher than that of the non-metastatic population. Yazdani et al. reported that more cancerous axillary lymph nodes increase the risk of BM (47). Colleoni et al. also found that there was the highest frequency of BM in patients who had four or more cancerous axillary lymph nodes (48). In addition, patients with higher-grade IDC had a higher risk of BM than those with grade I (Well differentiated) in this study. One reason for this may be that poorly differentiated tumor tissue was poorly demarcated from surrounding normal tissue, which meant it was more aggressive.

Additionally, it has been found that different breast subtypes showed different trends of BM. Previous study indicated that the incidence of BM was higher in patients with HR+/HER2- (Luminal A) and HR+/HER2+ (Luminal B) (49). In the current study, we found patients with HR+/HER2+ (Luminal B) were more likely to develop BM than those with other subtypes of BC.

Another risk factor for BM in our model was marital status. Unmarried patients had a higher risk of BM than those who were married. Zhao et al. and Gao et al. reported that marital status was a prognostic factor influencing the survival of metastatic BC patients (50, 51). Thus, we believed that the lifestyle habits and psychological factors of unmarried women may influence the distance metastasis of IDC patients.

As far as we know, this study was the first attempt to use ML algorithms to predict the risk of BM for IDC patients. Although some previous prediction models for BC based on the SEER database have been developed, it verified just from the SEER database and whether it can be used in different regions was not clear (26, 45). In addition, these studies only used nomogram as a visualization tool and did not provide a web predictor. External test of the model was important to validate the stability of the model for different regional populations. Therefore, in this study, we used data from the SEER database to build a prediction model based on the XGB algorithm and validated it with a cohort from China. Through our predictive model, we can predict the risk of BM in patients with IDC in the early

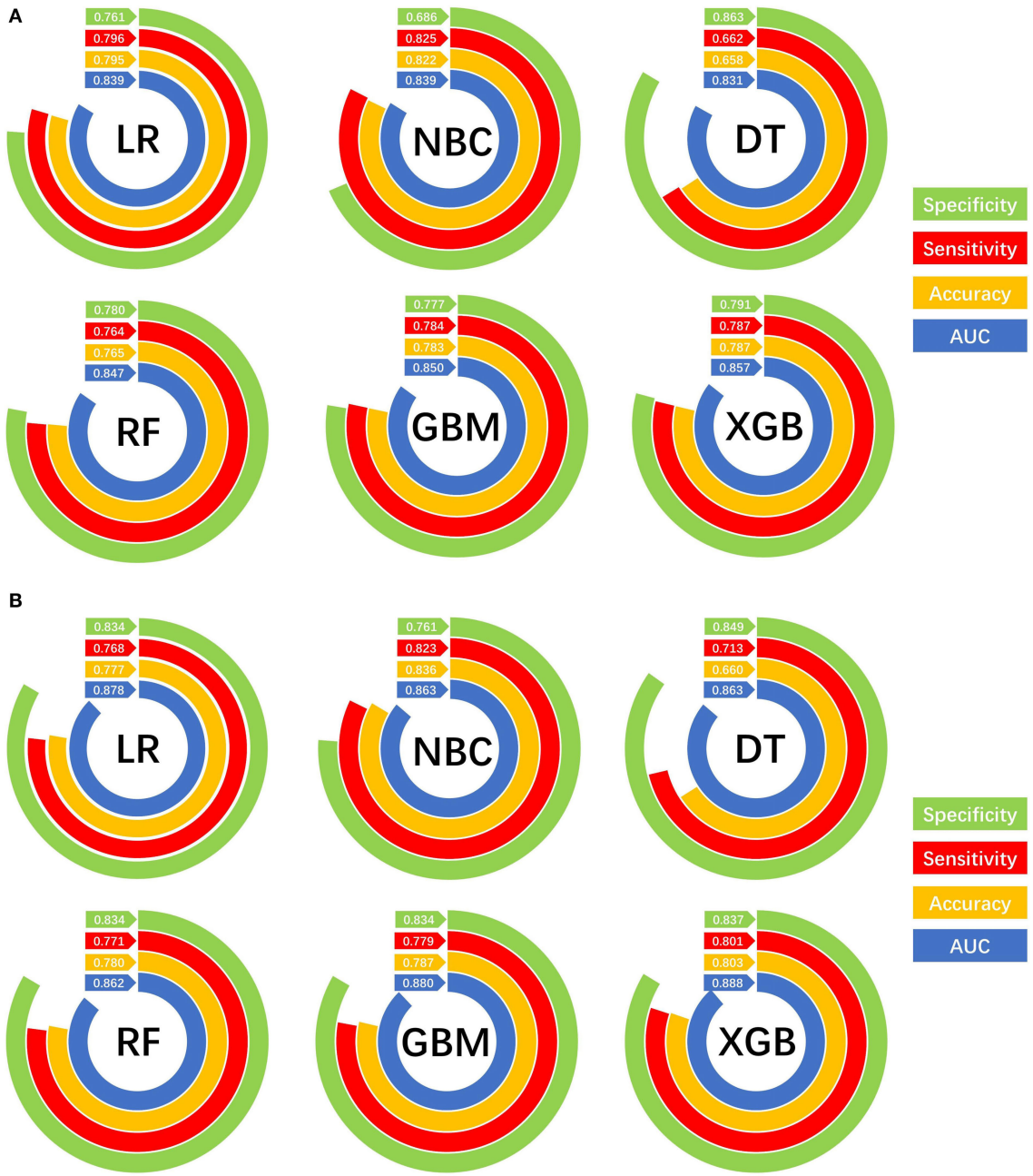


FIGURE 6 | Prediction performances of different machine learning models. **(A)** Internal validation of different machine learning models. **(B)** External validation of different machine learning models.

stage, indicate the related risks before the progression to the late stage, and accept the corresponding treatment regimen as soon as possible, which can significantly improve the prognosis of IDC patients. Furthermore, we developed a web predictor based on the model to predict the risk of BM in IDC patients. Clinicians can easily enter information about the patient's relevant variables into the model on the web page, and the model will calculate the patient's risk of developing BM. Nowadays, precision medicine has preceded four concepts: predictive, personalized, preventive and participatory. Due to the ML model we built, we can predict the risk of BM to a particular patient. The rapidity and accuracy of the prediction output allowed clinicians to make personalized decisions for their patients and could be used as a basis for clinicians to explain their decisions to patients and involved them in their treatment choices. From the clinician's point of view, substantial advances in ML had potential implications in clinical practice, including diagnosis, risk stratification and prognosis, treatment planning, and advances in precision medical methods

TABLE 4 | Comparison of prediction performances among different models for bone metastasis.

Models	Internal test				External test			
	AUC	Accuracy	Sensitivity (recall rate)	Specificity	AUC	Accuracy	Sensitivity (recall rate)	Specificity
LR	0.839	0.795	0.796	0.761	0.878	0.777	0.768	0.834
NBC	0.839	0.822	0.825	0.686	0.863	0.836	0.823	0.761
DT	0.831	0.658	0.662	0.836	0.863	0.660	0.713	0.849
RF	0.847	0.765	0.764	0.780	0.862	0.780	0.771	0.834
GBM	0.850	0.783	0.784	0.777	0.880	0.787	0.779	0.834
XGB	0.857	0.787	0.787	0.791	0.888	0.803	0.801	0.837

DT, Decision tree; LR, Logistic regression; GBM, Gradient Boosting Machine; NBC, Naive Bayes classification; RF, Random Forest; XGB, Extreme gradient boosting.

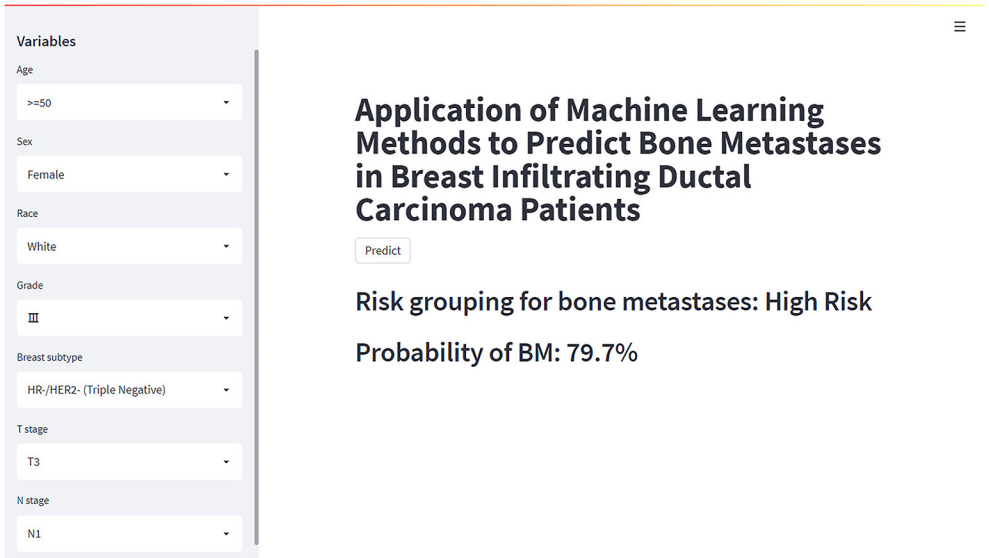


FIGURE 7 | The web calculator for predicting bone metastases in breast infiltrating duct carcinoma patients.

(52). Of course, they always had the final decision when it came to interpretation based on their domain expertise.

CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

Overall, this study used ML algorithms to construct and validate a clinical prediction model for predicting the risk of BM in patients with IDC based on large samples. Among all these algorithms, XGB performed the best. And we built an easy-to-use web calculator based on the XGB model, which can help physicians to individualize the diagnosis and treatment of BM in IDC patients.

Although our model achieved good results in prediction, there were still some limitations in it. First, it was a retrospective study, which needed to be further verified by prospective study. Second, only one external validation set was used to validate the model, and further efforts were required to validate the performance of the model in a more diverse population. Third, the SEER database just recorded the initial diagnosis of a patient, which meant that further information was lack and we were unable to access this information for further analysis.

For future work, we will focus on prospective and diverse population validation of the models to verify the performance and stability. These models are then expected to be integrated into applications that assist clinicians in medical decision-making. This can be a step toward a semi-autonomous diagnostic system that can assist clinicians in making individualized diagnoses of BM for IDC patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://seer.cancer.gov/.

ETHICS STATEMENT

We received permission to access the research data file in the SEER program from the National Cancer Institute, US. Approval

was waived by the Local Ethics Committee, as SEER data is publicly available and de-identified. This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanchang University, and cases from the First Affiliated Hospital of Nanchang University signed a written informed consent form.

AUTHOR CONTRIBUTIONS

W-CL and J-ML designed the study. W-CL, S-NW, W-LT, A-AL, and B-LS performed analysis and generated the figures and tables. W-CL, M-XL, and S-NW wrote the manuscript. Z-LL and J-ML critically reviewed the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by the Department of Science and Technology Program of Jiangxi Province, China (No. 20203BBG73045).

ACKNOWLEDGMENTS

We thank Mr. Wenxing Qian of Department of Computer Science, Beijing Jiaotong University for his assistance in algorithm and Ms. Yixin Lai of School of Pharmacy, Tongji Medical College, Huazhong University of Science and Technology for her assistance in literature search.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics (2021. *CA Cancer J Clin.* (2021) 71:7–33. doi: 10.3322/caac.21654
2. Pfeiffer RM, Webb-Vargas Y, Wheeler W, Gail MH. Proportion of US trends in breast cancer incidence attributable to long-term changes in risk factor distributions cancer. *Epidemiol Biomarkers Prev.* (2018) 27:1214–22. doi: 10.1158/1055-9965.EPI-18-0098
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
4. Molland J, Donnellan M, Janu N, Carmalt H, Kennedy C, Gillett D. Infiltrating lobular carcinoma—a comparison of diagnosis, management and outcome with infiltrating duct carcinoma. *Breast.* (2004) 13:389–96. doi: 10.1016/j.breast.2004.03.004
5. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA Cancer J Clin.* (2016) 66:31–42. doi: 10.3322/caac.21320
6. Zhang H, Zhu W, Biskup E, Yang W, Yang Z, Wang H, et al. Incidence, risk factors and prognostic characteristics of bone metastases and skeletal-related events (SREs) in breast cancer patients: a systematic review of the real world data. *J Bone Oncol.* (2018) 11:38–50. doi: 10.1016/j.jbo.2018.01.004
7. Chen WZ, Shen JF, Zhou Y, Chen XY, Liu JM, Liu ZL. Clinical characteristics and risk factors for developing bone metastases in patients with breast cancer. *Sci Rep.* (2017) 7:11325. doi: 10.1038/s41598-017-11700-4
8. Yamashiro H, Takada M, Nakatani E, Imai S, Yamauchi A, Tsuyuki S, et al. Prevalence and risk factors of bone metastasis and skeletal related events in patients with primary breast cancer in Japan. *Int J Clin Oncol.* (2014) 19:852–62. doi: 10.1007/s10147-013-0643-5
9. Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol.* (2010) 28:3271–7. doi: 10.1200/JCO.2009.25.9820
10. Kuchuk I, Hutton B, Moretto P, Ng T, Addison CL, Clemons M. Incidence, consequences and treatment of bone metastases in breast cancer patients-experience from a single cancer centre. *J Bone Oncol.* (2013) 2:137–44. doi: 10.1016/j.jbo.2013.09.001
11. Lüftner D, Lorusso V, Duran I, Hechmati G, Garzon-Rodriguez C, Ashcroft J, et al. Health resource utilization associated with skeletal-related events in patients with advanced breast cancer: results from a prospective, multinational observational study. *Springerplus.* (2014) 3:328. doi: 10.1186/2193-1801-3-328
12. Alonso-Betanzos A, Bolón-Canedo V. Big-data analysis, cluster analysis, and machine-learning approaches. *Adv Exp Med Biol.* (2018) 1065:607–26. doi: 10.1007/978-3-319-77932-4_37
13. Baştanlar Y, Ozuysal M. Introduction to machine learning. *Methods Mol Biol.* (2014) 1107:105–28. doi: 10.1007/978-1-62703-748-8_7
14. Li W, Hong T, Liu W, Dong S, Wang H, Tang ZR, et al. Development of a machine learning-based predictive model for lung metastasis in patients with ewing sarcoma. *Front Med (.* (2022) 9:807382. doi: 10.3389/fmed.2022.807382
15. Li W, Liu W, Hussain Memon F, Wang B, Xu C, Dong S, et al. An external-validated prediction model to predict lung metastasis among osteosarcoma: a multicenter analysis based on machine learning. *Comput Intell Neurosci.* (2022) 2022:2220527. doi: 10.1155/2022/2220527
16. Li W, Yafeng L, Liu W, Tang Z, Dong S, Lei M, et al. Machine learning-based prediction of lymph node metastasis among osteosarcoma patients. *Front Oncol.* (2022) 12:1418. doi: 10.3389/fonc.2022.797103
17. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics (.* (2010) 26:1340–7. doi: 10.1093/bioinformatics/btq134
18. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324
19. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining.* (2016). p. 785–94.
20. Qi Y. Random forest for bioinformatics. *Ensemble Mach Learn.* (2012) 307–23.
21. Tang J, Deng C, Huang G-B. Extreme learning machine for multilayer perceptron. *IEEE Trans Neural Netw Learn Syst.* (2015) 27:809–21. doi: 10.1109/TNNLS.2015.2424995
22. Sperandei S. Understanding logistic regression analysis. *Biochem Med.* (2014) 24:12–8. doi: 10.11613/BM.2014.003
23. Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. *J Chem Soc.* (2004) 18:275–85. doi: 10.1002/cem.873
24. Rish I. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Vol. 3.* (2001). p. 41–6.
25. Costa L, Badia X, Chow E, Lipton A, Wardley A. Impact of skeletal complications on patients' quality of life, mobility, and functional independence. *Support Care Cancer.* (2008) 16:879–89. doi: 10.1007/s00520-008-0418-0
26. Wu Q, Li J, Zhu S, Wu J, Chen C, Liu Q, et al. Breast cancer subtypes predict the preferential site of distant metastases: a SEER based study. *Oncotarget.* (2017) 8:27990–6. doi: 10.18632/oncotarget.15856
27. Fornetti J, Welm AL, Stewart SA. Understanding the bone in cancer metastasis. *J Bone Miner Res.* (2018) 33:2099–113. doi: 10.1002/jbmr.3618
28. Metzger-Filho O, Sun Z, Viale G, Price KN, Crivellari D, Snyder RD, et al. Patterns of Recurrence and outcome according to breast cancer subtypes in lymph node-negative disease: results from international breast cancer study group trials VIII and IX. *J Clin Oncol.* (2013) 31:3083–90. doi: 10.1200/JCO.2012.46.1574
29. Khan MU, Lee SUJ, Abbas S, Abbas A, Bashir AK. Detecting wake lock leaks in android apps using machine learning. *IEEE Access.* (2021) 9:125753–67. doi: 10.1109/ACCESS.2021.3110244
30. Akbar A, Ibrar M, Jan MA, Bashir AK, Wang L. SDN-enabled adaptive and reliable communication in IoT-Fog environment using machine learning

and multiobjective optimization. *IEEE Intern Thin J.* (2021) 8:3057–65. doi: 10.1109/JIOT.2020.3038768
31. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions. *Rev Literat.* (2019) 21:e12286. doi: 10.2196/12286
32. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med.* (2018) 284:603–19. doi: 10.1111/joim.12822
33. Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. *EBioMedicine.* (2019) 47:607–15. doi: 10.1016/j.ebiom.2019.08.027
34. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
35. Bhavsar KA, Abugabah A, Singla J, AlZubi AA, Bashir AK. A comprehensive review on medical diagnosis using machine learning. *Comput Mater Continua.* (2021) 67:1997. doi: 10.32604/cmc.2021.014943
36. Mohan S, Abugabah A, Kumar Singh S, Kashif Bashir A, Sanzogni L. An approach to forecast impact of Covid-19 using supervised machine learning model Software. *Pract Exp.* (2022) 52:824–40. doi: 10.1002/spe.2969
37. Iwendi C, Bashir AK, Peshkar A, Sujatha R, Chatterjee JM, Pasupuleti S, et al. COVID-19 patient health prediction using boosted random forest algorithm. *Front Public Health*. (2020) 8:357. doi: 10.3389/fpubh.2020.00357
38. Ayoub A, Mahboob K, Javed A-R, Rizwan M, Gadekallu T-R, Abidi M-H, et al. Classification and categorization of COVID-19 outbreak in Pakistan. *Comput Materi Continua*. (2021) 69 (:1253–69. doi: 10.32604/cmc.2021.015655
39. Bhavsar KA, Singla J, Al-Otaibi YD, Song O-Y, Zikria YB, Bashir AK. Medical diagnosis using machine learning: a statistical review. *Comput Mater Continua.* (2021) 67:107–25. doi: 10.32604/cmc.2021.014604
40. Liu WC Li ZQ, Luo ZW, Liao WJ, Liu ZL, Liu JM. Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer. *Cancer Med.* (2021) 10:2802–11. doi: 10.1002/cam4.3776
41. Liu W-C, Li M-X, Qian W-X, Luo Z-W, Liao W-J, Liu Z-L, et al. Application of machine learning techniques to predict bone metastasis in patients with prostate cancer. *Cancer Manag Res.* (2021) 13:8723. doi: 10.2147/CMAR.S330591
42. Huang Z, Hu C, Liu K, Yuan L, Li Y, Zhao C, et al. Risk factors, prognostic factors, and nomograms for bone metastasis in patients with newly diagnosed infiltrating duct carcinoma of the breast: a population-based study. *BMC Cancer.* (2020) 20:1–17. doi: 10.1186/s12885-020-07635-1
43. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. *Nat Rev Cancer.* (2005) 5:591–602. doi: 10.1038/nrc1670
44. DeSantis C, Jemal A, Ward E. Disparities in breast cancer prognostic factors by race, insurance status, and education. *Cancer Causes Control.* (2010) 21:1445–50. doi: 10.1007/s10552-010-9572-z
45. Chen MT, Sun HF, Zhao Y, Fu WY, Yang LP, Gao SP, et al. Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. *Sci Rep.* (2017) 7:9254. doi: 10.1038/s41598-017-10166-8
46. Wang R, Zhu Y, Liu X, Liao X, He J, Niu L. The Clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer. *BMC Cancer.* (2019) 19:1091. doi: 10.1186/s12885-019-6311-z
47. Yazdani A, Dorri S, Atashi A, Shirafkan H, Zabolinezhad H. Bone metastasis prognostic factors in breast cancer. *Breast Cancer.* (2019) 13:1178223419830978. doi: 10.1177/1178223419830978
48. Colleoni M, O'Neill A, Goldhirsch A, Gelber RD, Bonetti M, Thürlimann B, et al. Identifying breast cancer patients at high risk for bone metastases. *J Clin Oncol.* (2000) 18:3925–35. doi: 10.1200/JCO.2000.18.23.3925
49. Gong Y, Zhang J, Ji P, Ling H, Hu X, Shao ZM. Incidence proportions and prognosis of breast cancer patients with bone metastases at initial diagnosis. *Cancer Med.* (2018) 7:4156–69. doi: 10.1002/cam4.1668
50. Zhao W, Wu L, Zhao A, Zhang M, Tian Q, Shen Y, et al. A nomogram for predicting survival in patients with de novo metastatic breast cancer: a population-based study. *BMC Cancer.* (2020) 20:982. doi: 10.1186/s12885-020-07449-1
51. Gao T, Shao F. Risk factors and prognostic factors for inflammatory breast cancer with bone metastasis: a population-based study. *J Orthop Surg.* (2021) 29:23094990211000144. doi: 10.1177/23094990211000144
52. Rashidi HH, Tran N, Albahra S, Dang LT. Machine learning in health care and laboratory medicine: general overview of supervised learning and Auto-ML. *Int J Lab Hematol.* (2021) 43:15–22. doi: 10.1111/ijlh.13537

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Li, Wu, Tong, Li, Sun, Liu and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 12 July 2022
doi: 10.3389/fpubh.2022.935521

A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients Undergoing Surgery With Prostate Cancer: A Population-Based Study

Zhaoxia Zhang[1,2,3,4,5,6], *Chenghao Zhanghuang*[1,2,3,4,5,6], *Jinkui Wang*[1,2,3,4,5,6], *Tao Mi*[1,2,3,4,5,6], *Jiayan Liu*[1,2,3,4,5,6], *Xiaomao Tian*[1,2,3,4,5,6], *Liming Jin*[1,2,3,4,5,6] *and Dawei He*[1,2,3,4,5,6]*

[1] *Department of Urology, Chongqing Key Laboratory of Children Urogenital Development and Tissue Engineering, Chongqing, China,* [2] *Chongqing Key Laboratory of Pediatrics, Chongqing, China,* [3] *Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing, China,* [4] *National Clinical Research Center for Child Health and Disorders, Chongqing, China,* [5] *China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Chongqing, China,* [6] *Chongqing Higher Institution Engineering Research Center of Children's Medical Big Data Intelligent Application, Children's Hospital of Chongqing Medical University, Chongqing, China*

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Praise Young,
University of Eastern Finland, Finland
Joseph Henry Anajemba,
Hohai University, China
Michael Oyenuga,
Veritas University, Nigeria

***Correspondence:**
Dawei He
hedawei@hospital.cqmu.edu.cn

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *04 May 2022*
Accepted: *20 June 2022*
Published: *12 July 2022*

Citation:
Zhang Z, Zhanghuang C, Wang J, Mi T, Liu J, Tian X, Jin L and He D (2022) A Web-Based Prediction Model for Cancer-Specific Survival of Elderly Patients Undergoing Surgery With Prostate Cancer: A Population-Based Study. Front. Public Health 10:935521. doi: 10.3389/fpubh.2022.935521

Objective: Prostate cancer (PC) is the second leading cause of cancer death in men in the United States after lung cancer in global incidence. Elderly male patients over 65 years old account for more than 60% of PC patients, and the impact of surgical treatment on the prognosis of PC patients is controversial. Moreover, there are currently no predictive models that can predict the prognosis of elderly PC patients undergoing surgical treatment. Therefore, we aimed to construct a new nomogram to predict cancer-specific survival (CSS) in elderly PC patients undergoing surgical treatment.

Methods: Data for surgically treated PC patients aged 65 years and older were obtained from the Surveillance, Epidemiology, and End Results (SEER) database. Univariate and multivariate Cox regression models were used to identify independent risk factors for elderly PC patients undergoing surgical treatment. A nomogram of elderly PC patients undergoing surgical treatment was developed based on the multivariate Cox regression model. The consistency index (C-index), the area under the subject operating characteristic curve (AUC), and the calibration curve were used to test the accuracy and discrimination of the predictive model. Decision curve analysis (DCA) was used to examine the potential clinical value of this model.

Results: A total of 44,975 elderly PC patients undergoing surgery in 2010–2018 were randomly assigned to the training set (N = 31705) and validation set (N = 13270). the training set was used for nomogram development and the validation set was used for internal validation. Univariate and multivariate Cox regression model analysis showed that age, marriage, TNM stage, surgical style, chemotherapy, radiotherapy, Gleason score(GS), and prostate-specific antigen(PSA) were independent risk factors for CSS in elderly PC patients undergoing surgical treatment. The C index of the training set and validation indices are 0.911(95%CI: 0.899–0.923) and 0.913(95%CI:

0.893–0.933), respectively, indicating that the nomogram has a good discrimination ability. The AUC and the calibration curves also show good accuracy and discriminability.

Conclusions: To our knowledge, our nomogram is the first predictive model for elderly PC patients undergoing surgical treatment, filling the gap in current predictive models for this PC patient population. Our data comes from the SEER database, which is trustworthy and reliable. Moreover, our model has been internally validated in the validation set using the C-index,AUC and the and the calibration curve, showed that the model have good accuracy and reliability, which can help clinicians and patients make better clinical decision-making. Moreover, the DCA results show that our nomogram has a better potential clinical application value than the TNM staging system.

Keywords: nomogram, SEER, prostate cancer, surgery, elderly, CSS

BACKGROUND

The global incidence of prostate cancer (PC) is second only to lung cancer, and it is also the second leading cause of cancer death among men in the United States (1). PC is a heterogeneous disease with a variable natural history (2). With the introduction of serum prostate-specific antigen(PSA) testing in the late 1980s, the age at diagnosis of PC was nearly 10 years earlier, so surgeons can now provide more prompt surgical intervention early in tumor development without waiting for advanced stage (3).

PC has a variety of treatment options, including active surveillance (AS), surgical resection, radiotherapy (RT), chemotherapy, local ablation therapy (cryotherapy, high-intensity focused ultrasound), etc. Surgery is the primary treatment modality for PC, including radical prostatectomy (RP) and local tumor resection. However, while up to half of the patients with PC require only active surveillance at initial diagnosis, most men are willing to undergo aggressive local surgical treatment. Radical prostatectomy is associated with improved cancer-specific mortality(CSM), metastasis-free survival, and the need for palliative care, and surgical resection may be more beneficial for PC patients than radiotherapy (4). Xia et al. demonstrated that men undergoing radical prostatectomy had an increased life expectancy by 1.8 months compared to active surveillance (5). The above studies suggest that patients with PC treated with surgery may have a better prognosis. However, some studies suggest that surgery does not have much benefit in the prognosis of PC patients. The ProtecT trial reported the results of several extensive cohort studies suggesting that active surveillance could be safely used in carefully selected low-risk PC patients (6, 7). Elisabeth M et al. also found that surgical resection and radiotherapy did not benefit much from survival in patients with low-risk PC (4). We found that the current study is controversial regarding the impact of surgical treatment on the prognosis of PC patients. Based on this, we are eager to develop a model that can predict prognostic factors in PC patients undergoing surgery.

The traditional tumor-node-metastasis (TNM) cancer staging system of the American Joint Committee on Cancer (AJCC) has been used as a prognostic criterion for the vast majority of solid tumors. However, it does not include many important factors related to PC survival, such as age, prostate-specific antigen, Gleason score(GS), surgical, chemoradiotherapy, etc. In recent years, the nomogram prediction model has been considered one of the most accurate ways to predict tumors (8). Currently, nomograms based on the SEER database have been increasingly used to predict the prognosis of PC patients. Joao Ricardo Alves et al. constructed a nomogram predicting extracapsular extension in prostate cancer patients (9). In addition, researchers have developed nomograms that can predict lymph node metastases in patients with intermediate-risk prostate cancer, bone metastases in patients with newly diagnosed prostate cancer, survival in patients with prostate cancer spinal metastases, and survival in patients with metastatic prostate cancer undergoing radical prostatectomy (10–13). The incidence of PC is closely related to age. Data report that older men over 65 years old account for more than 60% of PC patients (14), and more than 90% of PC cancer-specific deaths (CSS) occur in this age group (15). Therefore, finding prognostic factors associated with CSS in surgically treated elderly PC patients has important clinical implications for reducing patient mortality. Unfortunately, however, to our knowledge, no nomogram has been developed for elderly PC patients undergoing surgical treatment. Therefore, our study aimed to use data from the SEER database to develop a nomogram for elderly PC patients undergoing surgical treatment that can accurately predict the relevant factors affecting patients' CSS to better assist clinicians and patients in making clinically aided decisions.

PATIENTS AND METHODS

Data Source and Data Extraction

Data from patients diagnosed with PC between 2010 and 2018 were extracted from the SEER database, and our study targeted patients aged 65 years and older and undergoing surgical treatment. The SEER database is a national cancer database covering approximately 30% of the population and containing 18 cancer medical centers. No ethical approval and patient informed consent is required because the SEER database data are publicly available and hidden patient information. The research methodology used in this study follows the research guidelines published in the SEER database.

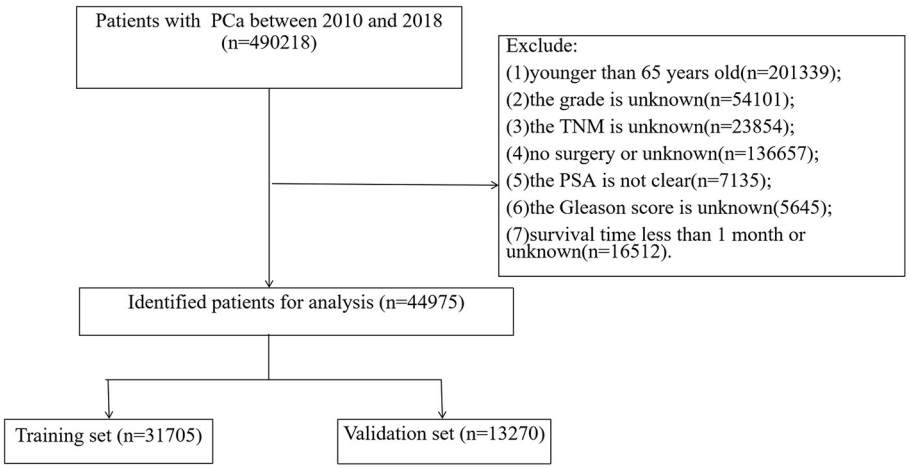


FIGURE 1 | Flowchart for inclusion and exclusion of elderly PC patients undergoing surgery treatment.

The variables in the SEER database include demographic characteristics [age, race, marital status), tumor grade (grade I, II, III, and IV)], TNM stage, surgical method (local tumor resection, radical prostatectomy), radiotherapy, chemotherapy, prostate-specific antigen, Gleason score, survival status, cause of death and survival time. Patient ethnicity was classified as white, black, and other types. Inclusion criteria: (1) patients aged 65 years and older, (2) a pathological diagnosis of PC, and (3) patients undergoing surgical treatment. Exclusion criteria: (1) patients younger than 65 years old, (2) tumor grade is unknown, (3) TNM stage is unknown, (4) no surgical treatment or surgical method is unknown, (5) prostate-specific antigen is unclear, (6) Gleason score is unknown, (7) survival time is < 1 month or survival time is unknown. The flowchart of patient inclusion and exclusion is shown in **Figure 1**.

Development and Validation of the Nomogram

All patients were randomized into the training set (70%) and the validation set (30%) for nomogram development and internal validation. Univariate and multivariate Cox proportional regression models were used to identify independent risk factors for patients in the training set. The nomogram was established based on the restivariate Cox regression analysis, and results were used to predict CSS at 3-, 5-, and 8-year in surgically treated elderly PC patients. In addition, the nomograms were calibrated at 3-, 5-, and 8-year, and these assessments used 1,000 autonomous samples. The accuracy and discrimination of the model were tested by the consistency index (c-index) and the area under the subject operating characteristic curve (AUC).

Clinical Application

The decision analysis curve (DCA) is used to assess the clinical value of the nomogram. We also calculated the risk for each patient from the nomogram. All patients were divided into high-risk and low-risk groups using the cutoff value of the subject operating characteristic curve (ROC). The Log-rank test and the Kaplan-Meier (K-M) curves examined the differences in survival between high-risk and low-risk patients. In addition, we analyzed differences between surgery and radiotherapy in high-risk and low-risk groups.

Statistical Analysis

Means and standard deviations were used to describe continuous variables such as age. Chi-square or non-parametric U tests were used for comparison between groups. Other classifications were described by frequency (%), and the differences between the groups were compared using the chi-square test. The Cox regression model analyzed the patient prognostic factors, and the patient survival differences were analyzed by the log-rank test and the K-M curve. Statistical analyses were performed using the R software version 4.1.0 and SPSS26.0.The R package includes "survival," "ggDCA," "DynNom," and "RMS" P values <0.05 were considered statistically significant.

RESULTS

Clinical Features

A total of 44,975 PC patients from 2010 to 2018 were included in our study from the SEER database who were all over 65 years old and were treated surgically, including the training set ($N = 31705$) and the validation set ($N = 13270$). The mean age of all patients was 70.0 ± 4.73 years, and most patients were white (82.7%) and married (76%). The tumor grade was mainly at grade II (42.2%) and grade III (48.9%). The T stage was dominated by T2 (48.6%), followed by T1 (22.6%) and T3 (26.5%), and minimal T4 (2.38%). All patients were mainly N0 (95.1%) and M0 (98.3%). Most patients underwent radical prostatectomy (83.2%), and fewer patients underwent local tumor resection (16.8%). Only 0.5% of the patients received chemotherapy. The majority of the patients (90.6%) did not receive radiotherapy. The Gleason score was mainly 7 (50.6%), and no significant difference between patients with Gleason scores$\leq$6 (24.2%) and Gleason score$\geq$8 (25.2%). 73.6% patients with prostate-specific antigen <10 ng/ml,

17.8% with prostate-specific antigen 10–20 ng / ml and only 8.58% with prostate-specific antigen> 20 ng/ml.There was no significant statistical deviation in the clinical characteristics of both groups, and the results are shown in **Table 1**.

COX Regression Analysis

A univariate Cox regression model was used to analyze and screen influencing factors associated with survival in the training set. The results showed that age, marriage, grade Grade III, TNM stage, surgical method, chemotherapy, radiotherapy, prostate-specific antigen, and Gleason score were the prognostic factors affecting patient survival. Then, a multivariate Cox regression analysis was used to screen for independent risk factors associated with patient survival. The results showed that age, marriage, TNM stage, surgical method, radiotherapy, chemotherapy, prostate-specific antigen, and Gleason score were independent risk factors affecting patient CSS. All of the results are shown in **Table 2**.

Development and Validation of the Nomograms

We constructed a feasible nomogram based on a multivariable Cox regression analysis model to predict CSS at 3-, 5-, and 8-year in elderly PC patients undergoing surgery (**Figure 2**). Age, TNM stage, surgical method, prostate-specific antigen, and Gleason score are the most significant factors influencing CSS in elderly PC patients undergoing surgical treatment. In addition, radiotherapy, chemotherapy, and marriage also have some effects.

We used internal cross-validation to validate the model's accuracy and discriminability. The C-index for the training and validation sets is 0.911(95%CI: 0.899–0.923) and 0.913(95%CI: 0.893–0.933, respectively), indicating that the prediction model of the nomogram has a good recognition ability. The calibration curves of the training and validation sets show that the predicted values of the nomogram are highly consistent with the actual observations (**Figure 3**), indicating that the nomogram has good accuracy. AUC was used to detect the discrimination of the nomogram, AUC at 3-,5-and 8-year was 92.9, 91.7, and 88.3 in the training set, and in the validation set, AUC at 3-,5-, and 8-year was 91.9,91.1 and 91.7, respectively (**Figure 4**).

Clinical Application of the Nomogram

DCA responds to the clinical value of the nomogram, and our results showed that it has good clinical potential value in both the training group and the validation set (**Figure 5**). Furthermore, we calculated the risk value for each patient based on the nomogram and calculated the optimal cutoff value using the ROC curve. Patients were classified into high-risk (total score 114.33) and low-risk (total score <114.33). The K-M curve showed that the survival rate of patients in the low-risk group was significantly higher than that in both the high-risk group and the high-risk group (**Figure 6**). The high-risk group of patients with 3-, 5-, and 8-year survival rates were 97.7%, 93.1%, and 88.6%, respectively. The low-risk group's 3-, 5- and 8-year survival rates were 99.8%, 99.6%, and 99.1%, respectively. Although most of the high-risk group patients underwent radical prostatectomy, the patients undergoing local tumor resection had higher survival rates. Almost all patients in the low-risk group underwent radical prostatectomy, so there was no significant difference in patient

TABLE 1 | Clinicopathological characteristics of elderly PC patients recieved surgery.

	All	Training cohort	Validation cohort	
	N = 44975	*N* = 31705	*N* = 13270	p
Age	70.0 (4.73)	70.0 (4.73)	69.9 (4.74)	0.847
Race:				0.469
White	37215 (82.7%)	26192 (82.6%)	11023 (83.1%)	
Black	4016 (8.93%)	2861 (9.02%)	1155 (8.70%)	
Other	3744 (8.32%)	2652 (8.36%)	1092 (8.23%)	
Marital:				0.385
No/Unknown	10813 (24.0%)	7659 (24.2%)	3154 (23.8%)	
Married	34162 (76.0%)	24046 (75.8%)	10116 (76.2%)	
Grade:				0.765
I	3236 (7.20%)	2281 (7.19%)	955 (7.20%)	
II	19001 (42.2%)	13402 (42.3%)	5599 (42.2%)	
III	22003 (48.9%)	15517 (48.9%)	6486 (48.9%)	
IV	735 (1.63%)	505 (1.59%)	230 (1.73%)	
T:				0.893
T1	10153 (22.6%)	7184 (22.7%)	2969 (22.4%)	
T2	21838 (48.6%)	15380 (48.5%)	6458 (48.7%)	
T3	11913 (26.5%)	8393 (26.5%)	3520 (26.5%)	
T4	1071 (2.38%)	748 (2.36%)	323 (2.43%)	
N:				0.538
N0	42781 (95.1%)	30145 (95.1%)	12636 (95.2%)	
N1	2194 (4.88%)	1560 (4.92%)	634 (4.78%)	
M:				0.040
M0	44201 (98.3%)	31133 (98.2%)	13068 (98.5%)	
M1	774 (1.72%)	572 (1.80%)	202 (1.52%)	
Surgery:				0.375
Local tumor excision	7543 (16.8%)	5350 (16.9%)	2193 (16.5%)	
Radical prostatectomy	37432 (83.2%)	26355 (83.1%)	11077 (83.5%)	
Chemotherapy:				0.514
No/unknown	44748 (99.5%)	31540 (99.5%)	13208 (99.5%)	
Yes	227 (0.50%)	165 (0.52%)	62 (0.47%)	
Radiation:				0.813
No/unknown	40756 (90.6%)	28738 (90.6%)	12018 (90.6%)	
Yes	4219 (9.38%)	2967 (9.36%)	1252 (9.43%)	
Gleason:				0.716
≤6	10901 (24.2%)	7666 (24.2%)	3235 (24.4%)	
7	22736 (50.6%)	16013 (50.5%)	6723 (50.7%)	
≥8	11338 (25.2%)	8026 (25.3%)	3312 (25.0%)	
PSA:				0.644
<10	33122 (73.6%)	23338 (73.6%)	9784 (73.7%)	
10–20	7994 (17.8%)	5622 (17.7%)	2372 (17.9%)	
>20	3859 (8.58%)	2745 (8.66%)	1114 (8.39%)	
CSS:				0.117
Dead	1203 (2.67%)	873 (2.75%)	330 (2.49%)	
Alive	43772 (97.3%)	30832 (97.2%)	12940 (97.5%)	
Survival months	45.6 (31.2)	45.4 (31.2)	46.0 (31.3)	0.115

TABLE 2 | Univariate and multivariate analyses of CSS in training cohort.

	Univariate			Multivariate		
	HR	**95%CI**	**P**	**HR**	**95%CI**	**P**
Age	1.18	1.17–1.19	<0.001	1.079	1.072–1.086	<0.001
Race						
White						
Black	1.1	0.87–1.39	0.43			
Other	0.88	0.68–1.13	0.307			
Marital						
No/unknown						
Married	0.55	0.48–0.63	<0.001	0.76	0.7–0.824	<0.001
Grade						
I						
II	0.86	0.5–1.48	0.586			
III	4.67	2.8–7.79	<0.001			
IV	2.65	0.6–11.6	0.197			
T						
T1						
T2	0.16	0.13–0.19	<0.001	0.856	0.762–0.961	0.009
T3	0.32	0.27–0.39	<0.001	1.042	0.899–1.208	0.585
T4	2.98	2.46–3.61	<0.001	1.581	1.343–1.86	<0.001
N						
N0						
N1	5.31	4.48–6.29	<0.001	1.393	1.211–1.603	<0.001
M						
M0						
M1	43.65	37.49–50.84	<0.001	2.597	2.243–3.008	<0.001
Surgery						
Local tumor excision						
Radical nephrectomy	0.09	0.08–0.1	<0.001	2.597	2.243–3.008	<0.001
Chemotherapy						
No/Unknown						
Yes	13.54	9.99–18.36	<0.001	1.723	1.294–2.296	<0.001
Radiation						
No/Unknown						
Yes	2.4	2.03–2.84	<0.001	0.797	0.709–0.897	<0.001
PSA						
<10						
10–20	2.28	1.9–2.73	<0.001	1.196	1.081–1.323	<0.001
>20	11.37	9.8–13.21	<0.001	1.356	1.211–1.519	<0.001
Gleason						
≤6						
7	1.97	1.44–2.69	<0.001	1.204	1.079–1.343	0.001
≥8	17.53	13.21–23.27	<0.001	2.048	1.822–2.302	<0.001

survival (**Figure 7**). We found that the survival rate of patients in the high-risk group receiving radiotherapy was lower than those who did not, considering that most older patients in the high-risk group had comorbidities and side effects of radiotherapy. In contrast, the survival rate of patients receiving and without radiotherapy in the low-risk group was not significantly different (**Figure 8**).

DISCUSSION

Although the traditional TNM stage is widely used in cancer management, multiple factors affect the prognosis clinically, and the TNM stage does not include these clinical factors. The prostate-specific antigen and Gleason score are critical factors for the prognosis of PC patients. Therefore, the most commonly used risk stratification scheme, D'Amico, combines clinical-stage, prostate-specific antigen, and Gleason score (16). Patients were divided into three risk levels: low risk (clinical stage T1-T2a, prostate-specific antigen≤10 ng/mL and Gleason score≤6), medium risk (T2b or <10 ≤prostate-specifc antigen 20 ng/mL or Gleason score 7), or high-risk disease (stage T2c or prostate-specific antigen> 20 ng/mL or Gleason score≥8). Although the D'Amico risk assessment protocol included prostate-specific antigen and Gleason score compared with the TNM stage, it still ignores the impact of age, ethnicity, marriage, surgical style, and chemoradiotherapy on the survival of PC patients.

The SEER database contains data from 18 medical centers analyzing prominent samples representative of populations from different regions. The SEER database contains patient demographic characteristics (age, race, marital status), tumor grade, TNM stage, treatment mode, and prognostic information, including survival status, cause of death, and survival time. The nomogram is a convenient and reliable statistical prediction tool that also uses multiple clinical variables different from the TNM stage to predict the prognosis of cancer patients (17). The nomogram can provide an individualized prediction of CSS through a comprehensive analysis of these clinicopathological factors. Nomogram meets the requirements of the integrated model. It promotes personalized medicine to facilitate clinicians' use for prognosis prediction using the SEER database. Previous virus studies have developed nomograms for predicting the OS and CSS of multiple databases. Moreover, it has shown promising clinical potential value in renal cancer, endometrial stromal sarcoma, glioma, nephroblastoma, and other cancers (18–20).

It is well known that the chance of genetic mutations inducing cancer increases with age. Studies have shown that age has a crucial role in the survival rate of various cancers (21). PC is no exception, and its incidence was significantly associated with age. PC was more frequent in older men, and the median age of diagnosis was 66 years (22). Multiple predictive models for the PC also showed that age is an independent risk factor for the prognosis of PC patients (15). Our results also show that age is an essential factor affecting patient prognosis. In addition, we found that age was linearly associated with poorer prognostic outcomes in elderly patients older than 65 years, with no significant cutoff value.

It has been reported in much literature that marriage benefits the prognosis of most cancer patients due to more social and financial support after marriage, as well as partner psychological comfort to patients (23). Our results show that marriage is also a protective factor for elderly PC patients undergoing surgery. Previous literature has reported the prevalence of PC among different ethnic groups, and the prognosis was also significantly associated with race (24). However, our study found that race was not an independent risk factor for elderly PC patients undergoing

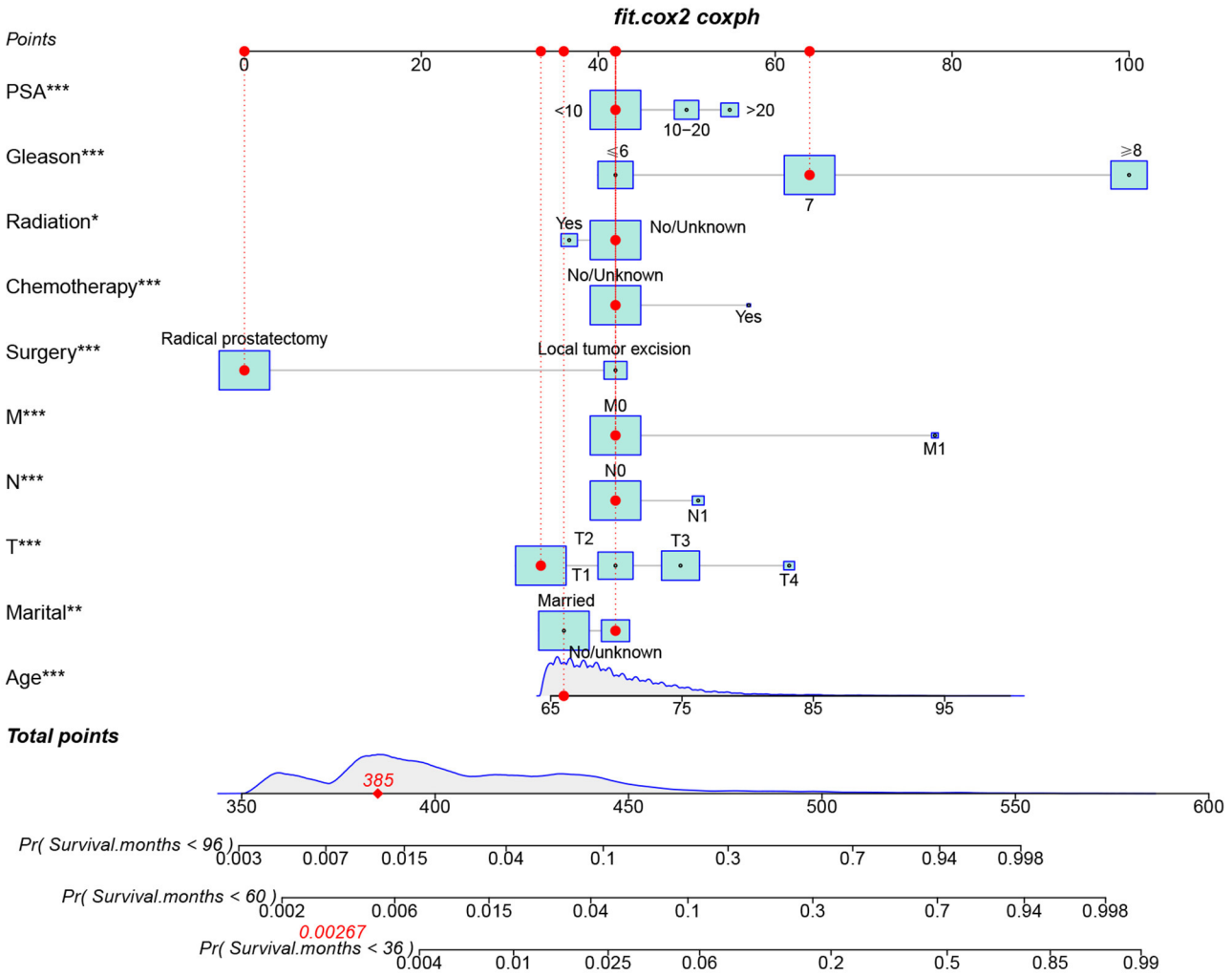


FIGURE 2 | The nomogram for predicting 3-, 5-, and 8-year CSS in elderly PC patients undergoing surgery treatment.

surgery. We considered that this might be related to differences in financial ability and medical level among different ethnic groups, leading to selection bias in treatment modalities.

Prostate-specific antigen screening began in the late 1980s, leading to a sharp increase in PC incidence over a short period (25). In 2012, the US Preventive Services Working Group (USPSTF) reported that prostate-specific antigen screening reduced the risk of PC death, and the USPSTF recommended including prostate-specific antigen in PC screening (26). The impact of screen-detected active treatment for PC on long-term survival compared to active surveillance is unknown but is associated with sexual dysfunction and dysuria (27). A secondary analysis of two randomized clinical trials by Martin et al. showed that increased prostate-specific antigen levels after treatment might be associated with a worse prognosis for locally advanced PC in men (28). The D'Amico risk assessment scheme divided prostate-specific antigen into three grades, prostate-specific antigen $\leq$10 ng/mL, 10–20 ng/mL, $>$ 20 ng/mL. Our prediction model also divided prostate-specific antigen according to this standard. The results showed that the higher the prostate-specific antigen expression level, the worse the patient prognosis, consistent with previous studies (29). Meanwhile, the Gleason score serves as an important tool for predicting the prognosis of PC patients. Our prediction model also divides it into three grades according to the D Amico protocol, Gleason score$\leq$6, Gleason score 7, and Gleason score$\geq$8. Our results showed that the Gleason score of most elderly PC patients undergoing surgical treatment was 7, the worst prognosis for Gleason score$\geq$8, and the best prognosis for Gleason score$\leq$6, which is also supported by previous studies (30, 31).

PC treatment mainly includes active surveillance, surgical resection, radiotherapy, chemotherapy, ADT, and other comprehensive treatment methods. Surgical resection mainly includes radical prostatectomy and local tumor resection. Radiotherapy and radical prostatectomy are evidence-based treatments for non-metastatic PC and can reduce PC mortality

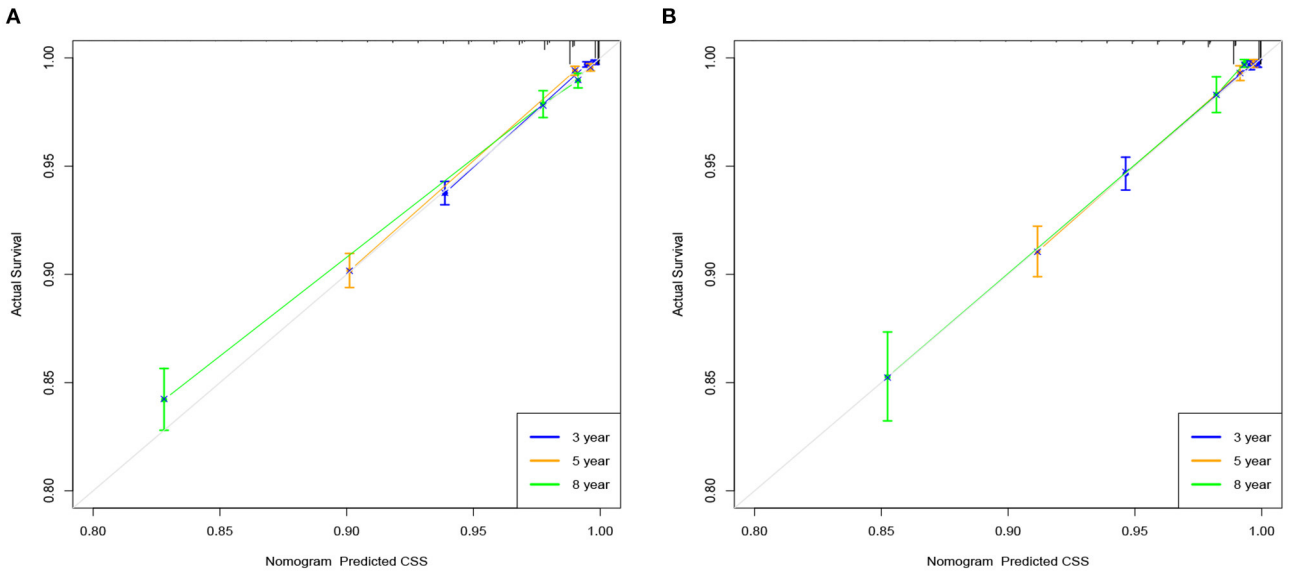


FIGURE 3 | Calibration curve of the nomogram in the training set **(A)** and validation set **(B)**. The horizontal axis is the predicted value in the nomogram, and the vertical axis is the observed value.

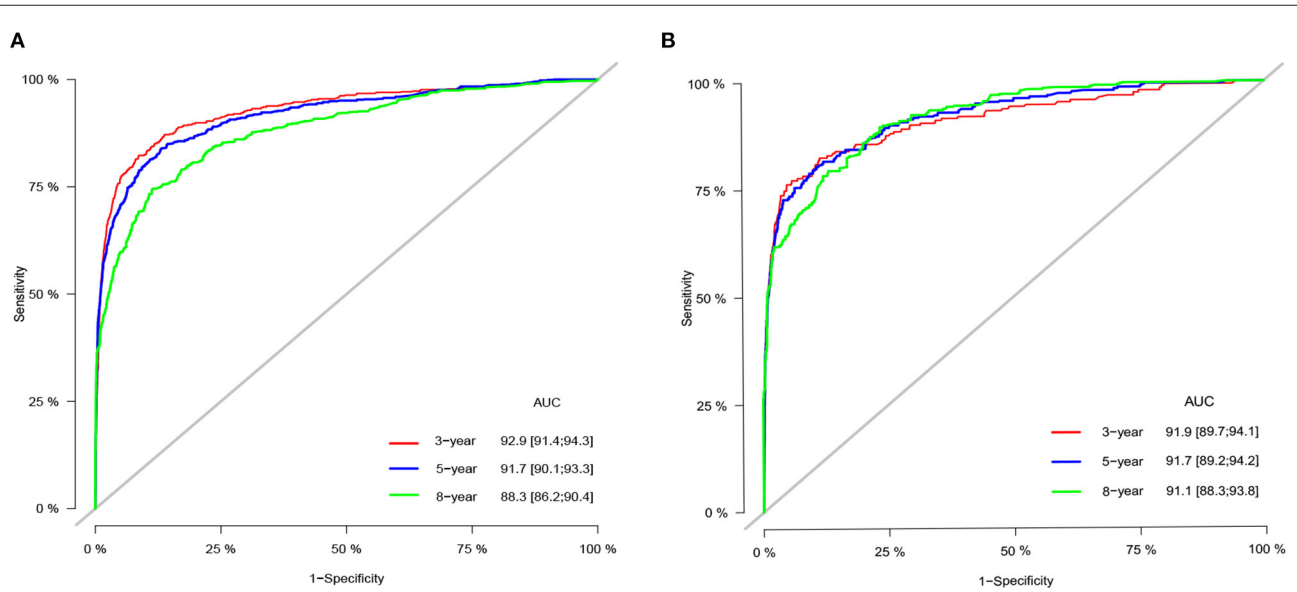


FIGURE 4 | AUC for predicting 3-, 5-, and 8-year CSS in the training set **(A)** and validation set **(B)**.

compared with non-therapeutic randomized clinical trials (32). A ProtecT randomized controlled trial by David E Neal et al. showed that 90% of localized patients did not die of PC within 10 years. Whether through active surveillance, surgery, or radiotherapy, active surveillance had fewer side effects on sexual and bladder function but a greater risk of tumor spread (33). The study by Freddie C Hamdy et al. confirmed that surgery and radiotherapy were associated with a lower incidence of disease progression and metastasis than active surveillance (34). A randomized trial conducted by Lars Holmberg et al. showed that radical prostatectomy significantly reduced CSS in early-stage PC patients but showed no significant difference between surgery and observation waiting for OS (40). Anna Bill-Axelson et al. confirmed that radical prostatectomy was associated with reduced mortality in PC, but adjuvant local or systemic therapy may be more beneficial in men with extracapsular tumor growth; they also concluded that radical prostatectomy showed no benefit in

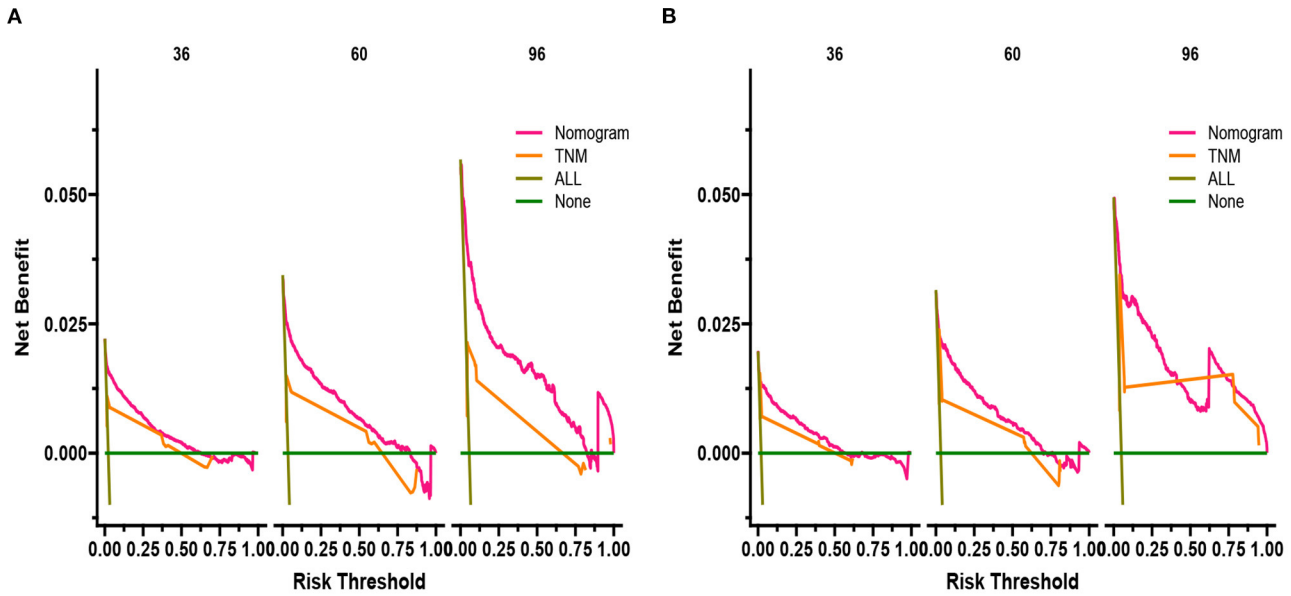


FIGURE 5 | DCA of the nomogram in the training set **(A)** and the validation set **(B)**. The Y-axis represents a net benefit, and the X-axis represents threshold probability. The green line means no patients died, and the dark green line means all patients died. When the threshold probability is between 0% and 100%, the net benefit of the model exceeds all deaths or none.

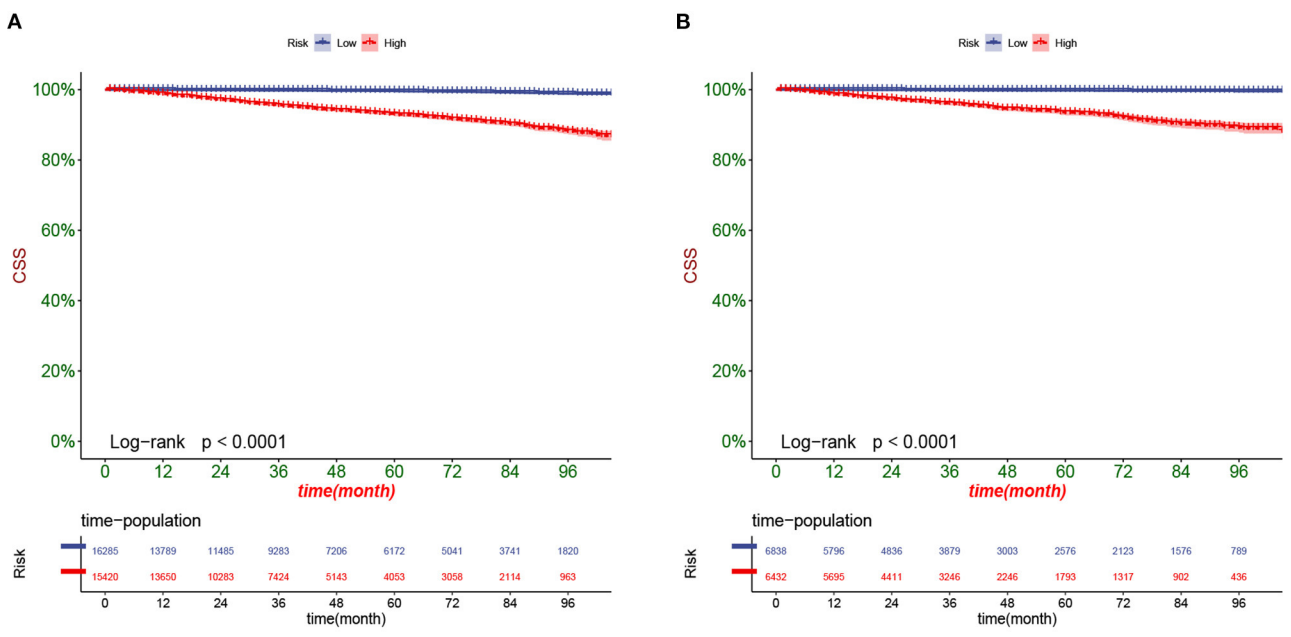


FIGURE 6 | Kaplan-Meier curves of patients in the low-risk and high-risk groups in the training set **(A)** and validation set **(B)**.

elderly PC patients over 65 years (35). After nearly 20 years of follow-up of men with localized PC, Timothy J Wilt et al. found that surgery was not associated with significantly reduced mortality and a higher frequency of adverse events than active surveillance (36). The above studies show that different treatment modalities' outcomes differ significantly for different types of PC patients. Our results showed that radical prostatectomy is associated with a better prognosis than local tumor resection; postoperative radiotherapy achieves a better prognosis than patients without radiotherapy. However, our nomogram shows that chemotherapy is associated with a worse prognosis. We consider why patients receiving chemotherapy often relapsed or

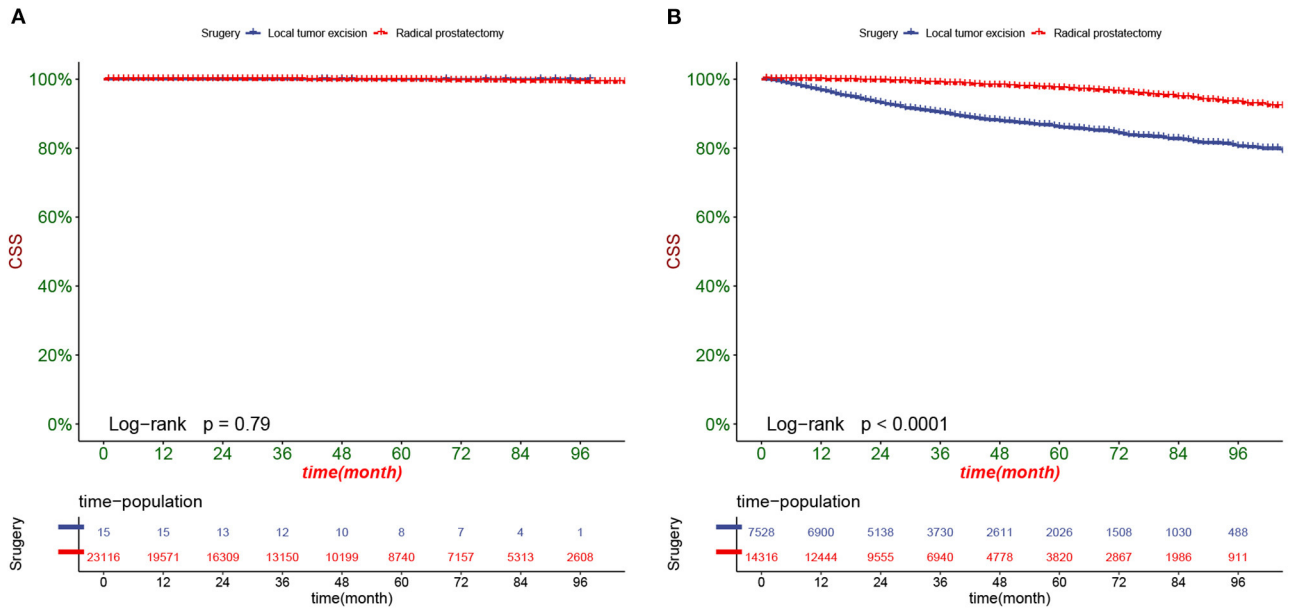


FIGURE 7 | Kaplan-Meier curves of patients with different surgery in the low-risk group **(A)** and high-risk group **(B)**.

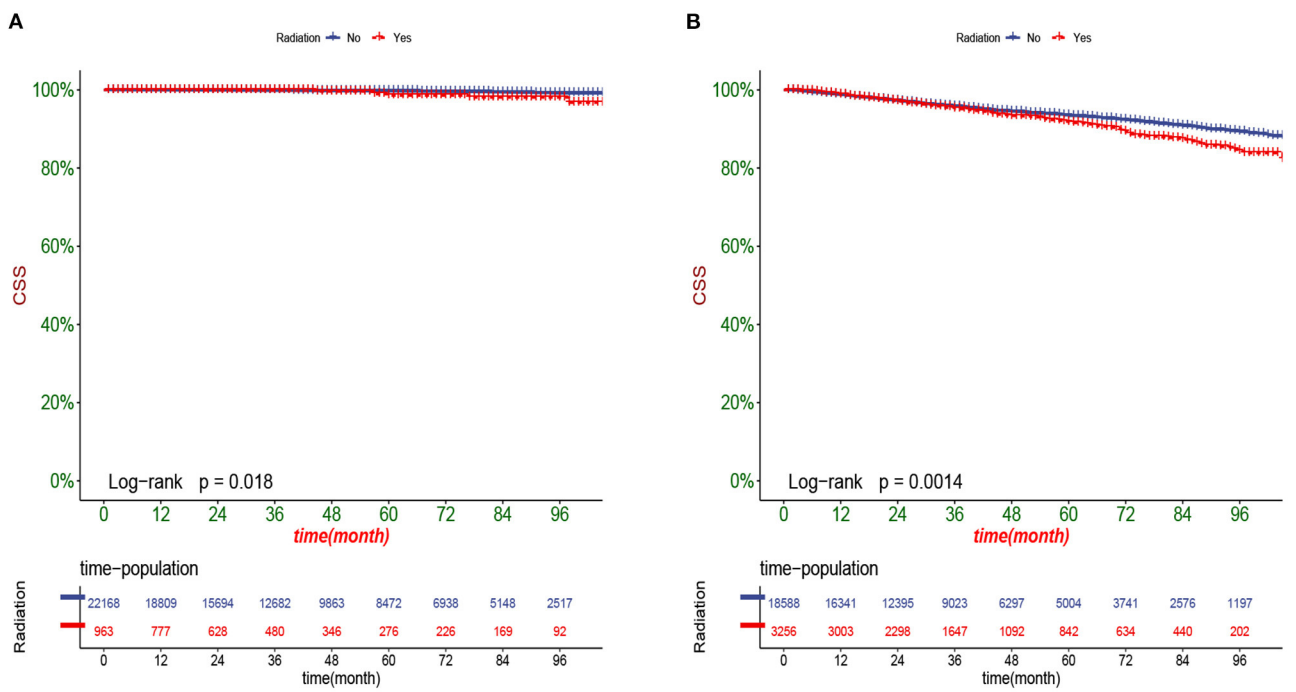


FIGURE 8 | Kaplan-Meier curves of patients with or without radiotherapy in the low-risk group **(A)** and high-risk group **(B)**.

have distant metastasis and have a worse prognosis. At the same time, the small number of people receiving chemotherapy results in biased results.

The nomogram based on the SEER database has good accuracy, but this study has some shortcomings. First, there is a lack of essential information, such as ADT, which is one of the primary treatment modalities for PC; however, there is a lack of data related to ADT in the SEER database, so our model also lacks the relationship between ADT and prognosis. Moreover, PSA is an important indicator related to prognosis for PC patients.

However, it was not included in the SEER database until 2010, so we can only choose data after 2010 to build a predictive model, resulting in a smaller sample size for us to choose. Second, our nomogram could not include all related prognostic variables such as BMI, smoking, alcohol consumption, Etc. leading to some limitations in our results. Finally, our study was retrospective, which may lead to an unavoidable risk of selection bias. Although our results have some limitations, our nomogram still includes key prognostic variables and has been internally validated to show good accuracy, so the results are not significantly biased. Second, our nomogram shows good potential clinical application value, which can help doctors and patients to make clinical decision-making.

CONCLUSION

Our study explored the clinicopathological factors affecting CSS in elderly PC patients undergoing surgical treatment. We found that age, marriage, prostate-specific antigen, Gleason score, surgical mode, chemoradiotherapy, and TNM stage were independent risk factors affecting CSS in patients. We established a nomogram that could predict the CSS in elderly PC patients undergoing surgical treatment. The model has been internally validated with good accuracy and reliability. It showed better potential clinical application value than the TNM staging system; it can help clinicians and patients make clinical decisions. However, our study is retrospective, which may lead to an unavoidable risk of selection bias. Future prospective studies with large and multicenter samples are needed to validate the nomogram further.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

ZZ and CZ designed the study. ZZ, JW, JL, LJ, and XT collected and analyzed the data. ZZ drafted the initial manuscript. CZ, TM, JL, and DH revised the article critically. CZ, JL, and DH reviewed and edited the article. All authors approved the final manuscript.

FUNDING

Special Key Project of Chongqing Technology Innovation and Application Development (No. Cstc2019jscx-tjsbX0003), Yunnan Education Department of Science Research Fund (No. 2020 J0228), Kunming City Health Science and Technology Talent 1000 training Project (No. 2020- SW (Reserve)-112), Kunming Health and Health Commission Health Research Project (No. 2020-0201-001), and Kunming Medical Joint Project of Yunnan Science and Technology Department (No. 202001 AY070001-271).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.935521/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* (2022) 72:7–33. doi: 10.3322/caac.21708
2. Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, et al. Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol.* (2021) 18:79–92. doi: 10.1038/s41585-020-00400-w
3. Bruinsma SM, Roobol MJ, Carroll PR, Klotz L, Pickles T, Moore CM, et al. Expert consensus document: semantics in active surveillance for men with localized prostate cancer - results of a modified Delphi consensus procedure. *Nat Rev Urol.* (2017) 14:312–22. doi: 10.1038/nrurol.2017.26
4. Sebesta EM, Anderson CB. The surgical management of prostate cancer. *Semin Oncol.* (2017) 44:347–57. doi: 10.1053/j.seminoncol.2018.01.003
5. Xia J, Trock BJ, Cooperberg MR, Gulati R, Zeliadt SB, Gore JL, et al. Prostate cancer mortality following active surveillance versus immediate radical prostatectomy. *Clin Cancer Res.* (2012) 18:5471–8. doi: 10.1158/1078-0432.CCR-12-1502
6. Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. *J Clin Oncol.* (2015) 33:272–7. doi: 10.1200/JCO.2014.55.1192
7. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. *Eur Urol.* (2013) 63:597–603. doi: 10.1016/j.eururo.2012.11.005
8. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7
9. Alves JR, Muglia VF, Lucchesi FR, Faria R, Alcantara-Quispe C, Vazquez VL, et al. Independent external validation of nomogram to predict extracapsular extension in patients with prostate cancer. *Eur Radiol.* (2020) 30:5004–10. doi: 10.1007/s00330-020-06839-0
10. Peilleron N, Seigneurin A, Herault C, Verry C, Bolla M, Rambeaud JJ, et al. External evaluation of the Briganti nomogram to predict lymph node metastases in intermediate-risk prostate cancer patients. *World J Urol.* (2021) 39:1489–97. doi: 10.1007/s00345-020-03322-9
11. Zhang W, Mao N, Wang Y, Xie H, Duan S, Zhang X, et al. A Radiomics nomogram for predicting bone metastasis in newly diagnosed prostate cancer patients. *Eur J Radiol.* (2020) 128:109020. doi: 10.1016/j.ejrad.2020.109020
12. Y. Liu, L. Li, D. Jiang, M. Yang, X. Gao, K. Lv, et al. A Novel Nomogram for Survival Prediction of Patients with Spinal Metastasis From Prostate Cancer. *Spine (Phila Pa 1976).* (2021) 46 E364–73. doi: 10.1097/BRS.0000000000003888
13. Wu K, Tang Y, Shao Y, Li X. Nomogram predicting survival to assist decision-making of radical prostatectomy in patients with metastatic prostate cancer. *Transl Androl Urol.* (2021) 10:879–87. doi: 10.21037/tau-20-1166
14. Meltzer D, Egleston B, Abdalla I. Patterns of prostate cancer treatment by clinical stage and age. *Am J Public Health.* (2001) 91:126–8. doi: 10.2105/AJPH.91.1.126
15. Siegel DA, O'Neil ME, Richards TB, Dowling NF, Weir HK. Prostate cancer incidence and survival, by stage and race/ethnicity - United States, 2001-2017. *MMWR Morb Mortal Wkly Rep.* (2020) 69:1473–80. doi: 10.15585/mmwr.mm6941a1
16. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam

radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA.* (1998) 280:969–74. doi: 10.1001/jama.280.11.969
17. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* (2008) 26:1364–70. doi: 10.1200/JCO.2007.12.9791
18. Zhanghuang C, Wang J, Zhang Z, Jin L, Tan X, Mi T, et al. A web-based prediction model for cancer-specific survival of elderly patients with clear cell renal cell carcinoma: a population-based study. *Front Public Health.* (2021) 9:833970. doi: 10.3389/fpubh.2021.833970
19. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. *Cancer Commun (Lond).* (2020) 40:301–12. doi: 10.1002/cac2.12067
20. Pan Z, You H, Bu Q, Feng X, Zhao F, Li Y, et al. Development and validation of a nomogram for predicting cancer-specific survival in patients with Wilms' tumor. *J Cancer.* (2019) 10:5299–305. doi: 10.7150/jca.32741
21. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, et al. Renal cell carcinoma: ESMO clinical practice guidelines for diagnosis, treatment and follow-updagger. *Ann Oncol.* (2019) 30:706–20. doi: 10.1093/annonc/mdz056
22. Droz JP, Albrand G, Gillessen S, Hughes S, Mottet N, Oudard S, et al. Management of prostate cancer in elderly patients: recommendations of a task force of the international society of geriatric oncology. *Eur Urol.* (2017) 72:521–31. doi: 10.1016/j.eururo.2016.12.025
23. Chen ZH, Yang KB, Zhang YZ, Wu CF, Wen DW, Lv JW, et al. Assessment of modifiable factors for the association of marital status with cancer-specific survival. *JAMA Netw Open.* (2021) 4:e2111813. doi: 10.1001/jamanetworkopen.2021.11813
24. R.J. Thorpe, Jr., M.A. Bruce, D.L. Howard, and T.A. LaVeist, Race differences in mobility status among prostate cancer survivors: The role of socioeconomic status. *Adv Cancer Res.* (2020) 146:103–14. doi: 10.1016/bs.acr.2020.01.006
25. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *J Natl Cancer Inst.* (2009) 101:1325–9. doi: 10.1093/jnci/djp278
26. Force USPST, Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, et al. Screening for prostate cancer: US preventive services task force recommendation statement. *JAMA.* (2018) 319:1901–13. doi: 10.1001/jama.2018.3710
27. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-specific antigen-based screening for prostate cancer: evidence report and systematic review for the us preventive services task force. *JAMA.* (2018) 319:1914–31. doi: 10.1001/jama.2018.3712
28. King MT, Chen MH, Collette L, Neven A, Bolla M, D'Amico AV. Association of increased prostate-specific antigen levels after treatment and mortality in men with locally advanced vs localized prostate cancer: a secondary analysis of 2 randomized clinical trials. *JAMA Netw Open.* (2021) 4:e2111092. doi: 10.1001/jamanetworkopen.2021.11092
29. Zhu X, Gou X, Zhou M. Nomograms predict survival advantages of Gleason Score 3+4 Over 4+3 for prostate cancer: a SEER-based study. *Front Oncol.* (2019) 9:646. doi: 10.3389/fonc.2019.00646
30. Ham WS, Chalfin HJ, Feng Z, Trock BJ, Epstein JI, Cheung C, et al. New prostate cancer grading system predicts long-term survival following surgery for gleason score 8-10 prostate cancer. *Eur Urol.* (2017) 71:907–12. doi: 10.1016/j.eururo.2016.11.006
31. Pinthus JH, Witkos M, Fleshner NE, Sweet J, Evans A, Jewett MA, et al. Prostate cancers scored as Gleason 6 on prostate biopsy are frequently Gleason 7 tumors at radical prostatectomy: implication on outcome. J Urol. (2006) 176:979–84. doi: 10.1016/j.juro.2006.04.102
32. Holmberg L, Bill-Axelson A, Helgesen F, Salo JO, Folmerz P, Haggman M, et al. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *N Engl J Med.* (2002) 347:781–9. doi: 10.1056/NEJMoa012794
33. Neal DE, Metcalfe C, Donovan JL, Lane JA, Davis M, Young GJ, et al. Ten-year mortality, disease progression, and treatment-related side effects in men with localised prostate cancer from the protect randomised controlled trial according to treatment received. *Eur Urol.* (2020) 77:320–30.
34. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med.* (2016) 375:1415–24. doi: 10.1056/NEJMoa1606220
35. Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med.* (2011) 364:1708–17. doi: 10.1056/NEJMoa1011967
36. Wilt TJ, Jones KM, Barry MJ, Andriole GL, Culkin D, Wheeler T, et al. Follow-up of prostatectomy versus observation for early prostate cancer. *N Engl J Med.* (2017) 377:132–42. doi: 10.1056/NEJMoa1615869

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Zhanghuang, Wang, Mi, Liu, Tian, Jin and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 27 July 2022
DOI 10.3389/fpubh.2022.946833

Check for updates

OPEN ACCESS

EDITED BY
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

REVIEWED BY
Liu Bingfeng,
Jingdezhen Ceramic University, China
Lin Haifeng,
McMaster University, Canada
Hui Feng,
Jiangsu Vocational College of Finance and Economics, China

*CORRESPONDENCE
Lingyu Sun
lingyusun115@csu.edu.cn

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 18 May 2022
ACCEPTED 22 June 2022
PUBLISHED 27 July 2022

CITATION
Zhuang L, Dai M, Zhou Y and Sun L (2022) Intelligent automatic sleep staging model based on CNN and LSTM. *Front. Public Health* 10:946833. doi: 10.3389/fpubh.2022.946833

COPYRIGHT
© 2022 Zhuang, Dai, Zhou and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Intelligent automatic sleep staging model based on CNN and LSTM

Lan Zhuang[1], Minhui Dai[2,3], Yi Zhou[3] and Lingyu Sun[2,3]*

[1]Staff Hospital, Central South University, Changsha, China, [2]Teaching and Research Section of Clinical Nursing, Xiangya Hospital of Central South University, Changsha, China, [3]Department of Ophthalmology, Xiangya Hospital, Central South University, Changsha, China

Since electroencephalogram (EEG) is a significant basis to treat and diagnose somnipathy, sleep electroencephalogram automatic staging methods play important role in the treatment and diagnosis of sleep disorders. Due to the characteristics of weak signals, EEG needs accurate and efficient algorithms to extract feature information before applying it in the sleep stages. Conventional feature extraction methods have low efficiency and are difficult to meet the time validity of fast staging. In addition, it can easily lead to the omission of key features owing to insufficient a priori knowledge. Deep learning networks, such as convolutional neural networks (CNNs), have powerful processing capabilities in data analysis and data mining. In this study, a deep learning network is introduced into the study of the sleep stage. In this study, the feature fusion method is presented, and long-term and short-term memory (LSTM) is selected as the classification network to improve the accuracy of sleep stage recognition. First, based on EEG and deep learning network, an automatic sleep phase method based on a multi-channel EGG is proposed. Second, CNN-LSTM is used to monitor EEG and EOG samples during sleep. In addition, without any signal preprocessing or feature extraction, data expansion (DA) can be realized for unbalanced data, and special data and non-general data can be deleted. Finally, the MIT-BIH dataset is used to train and evaluate the proposed model. The experimental results show that the EEG-based sleep phase method proposed in this paper provides an effective method for the diagnosis and treatment of sleep disorders, and hence has a practical application value.

KEYWORDS

EEG signal, convolutional neural network, long-term and short-term memory, sleep stage, multichannel, feature fusion

Introduction

People spend one-third of their time sleeping every day, and sleep is an important and necessary physiological process of the human body. Sleep quality will seriously affect people's physical and mental health. The acceleration of the pace of life in modern society and the increase of life pressure have also affected people's sleep quality, causing many people to suffer from sleep diseases (1). Clinical medicine analyzes and judges sleep quality by monitoring patients' sleep status, so as to achieve the purpose of treating or preventing sleep diseases. The collection of sleep data is the basis of sleep quality analysis.

At present, the research on automatic sleep staging has made some progress. The sleep staging process requires professionals to receive a lot of training and accumulate on-site experience, which is a complex and cumbersome process. Therefore, it is necessary to combine sleep with artificial intelligence to automatically and effectively evaluate sleep stages. Sleep is the most important physiological process related to human health. With the acceleration of people's pace of life and the increase in stress, the incidence rate of sleep disorders is also increasing. At present, patients tend to be younger. Therefore, the research on sleep has attracted extensive attention from scholars. From a clinical point of view, sleep staging is an important basis for monitoring sleep quality and detecting sleep-related diseases. Nowadays, people all over the world have difficulty sleeping or even insomnia, and more and more diseases are caused by sleep problems. Studies have shown that sleep quality plays a very important role in the regeneration of memory neurons. Poor sleep quality or insufficient sleep will damage memory to some extent. In addition, some studies have proved that sleep disorder has been identified as one of the earliest potential biomarkers for the progression of Alzheimer's disease, and monitoring sleep changes is helpful for early intervention of dementia (2). Therefore, how to correctly stage sleep is an important basis for monitoring sleep quality and diagnosing sleep-related diseases. It has important practical significance and application value to improve sleep quality or diagnose sleep disorders through artificial intelligence. The convolutional neural network is a feedforward neural network, including an input layer, an output layer, convolution layer, and pool layer (3). Its special network structure can effectively reduce the complexity of the neural network. It has achieved great success in many fields, particularly in image recognition. At present, CNN has been successfully applied to many specific projects, such as original EEG data, target recognition, and handwritten character recognition. It also occupies a place in many fields, such as analyzing visual images, speech recognition, and predicting human traffic (4). As we all know, CNN has the characteristics of shared weight structure and transformation invariance. CNN consists of input and output layers and multiple hidden layers. The hidden layer can be a convolution layer, a pooling layer, or a full connection layer. Compared with other traditional algorithms, CNN uses relatively less preprocessing. This means that CNN can learn from the manually designed filter in the traditional algorithm. Its main advantage is that it can replace prior knowledge and artificial design feature engineering (5). Therefore, this study aims to study a new automatic sleep classification method based on EEG signals, including signal preprocessing, feature extraction, and classification and recognition. The main research content of this paper is to realize the automatic segmentation of sleep by using signal processing method and pattern recognition theory, and establish an automatic sleep staging system with high accuracy and fast calculation speed to provide more effective and convenient sleep staging methods for medical researchers, and reduce the workload of clinical diagnosis.

The specific contributions of this study include the following:

(1) Based on EEG signals and a deep learning network, this paper proposes an automatic sleep staging method based on multi-channel EEG, and CNN-LSTM is used for supervised learning of sleep staging on original EEG and EOG samples.
(2) The EEG signals are filtered and extracted, respectively, and the features are extracted by CNN and LSTM in parallel. The two features are fused for classification.
(3) By introducing an electro-oculogram and fusing the modal information between EEG and EOG, the classification accuracy of each stage is improved.
(4) When decomposing EEG signal and using the method proposed in this paper and the characteristics of rhythm wave, relative energy, and complexity for automatic sleep staging, by using the information complementarity between different modes, the calculation time of feature extraction can be reduced, the error can be reduced, and the accuracy of automatic staging can be improved.

The remainder of this paper is organized as follows. Section Related work discusses related work, followed by the composition of brain waves in EEG and signal detection in Section EEG signal. Sleep staging criteria and specific sleep stages are discussed in Section Sleep stage. Section CNN-LSTM sleep stages algorithms presents CNN-LSTM sleep stages algorithms, Section Contrast experiment and result analysis details the contrast experiment and result analysis, and Section Conclusion and future work provides a conclusion with a summary and future research direction.

Related work

Sleep experts can determine the quality of sleep by recording electrical activity through sensors connected to different parts of the body. A set of signals collected by these sensors is called polysomnography (PSG). A series of physiological signals were recorded, including EEG, EOG, EMG, and ECG. EEG signals often contain a lot of information about physiological activities (6). It is an extremely complex bioelectrical signal. It provides important analytical reference information for neurology, medicine, and other disciplines. It can accurately reflect the activity of the brain and is gradually widely used in sleep research. Each person has different characteristics and amplitudes of EEG signals in different sleep states, which reflects the different and complex functions of the brain in different stages (7). The classification of different characteristics of different sleep stages is called sleep stage. It is not only

the standard for judging sleep quality, but also the basis for diagnosing some brain diseases. It has important clinical significance and broad application prospects for the treatment of sleep disorders (8). The early classification standard of sleep stage is the R&K standard, which was proposed by Rechtchaffen and Kales in 1968. The human sleep process is divided into three types: wake, non-rapid eye movement sleep (NREM), and rapid eye movement sleep (REM) (9). According to expert standards, polysomnography (PSG) is divided into stages every 30 s. Professional sleep data evaluation can only be carried out in the hospital, and then experts judge each paragraph according to the classification criteria through visual observation. It is very cumbersome, time-consuming, and prone to subjective errors (10). An 8-h PSG data requires professional doctors to spend 2–4 h on sleep staging, which is very inefficient. Therefore, the development of automatic sleep staging is of great significance. Automatic staging is the research of feature extraction and classification of EEG data collected by using signal processing, pattern recognition, and computer technology. Compared with manual staging, it saves time and effort. It can not only improve the efficiency of data judgment and classification, but also reduce the misjudgment caused by doctors' subjective factors in the process of staging (11, 12).

In order to avoid the misjudgment of subjective consciousness caused by the manual recognition of the sleep stage by experts, more and more scholars are committed to the research of automatic sleep staging. Sleep data staging belongs to the category of pattern recognition (13). In recent years, with the rapid development of pattern recognition methods, a large number of classification methods have been found and used, accompanied by continuous changes and improvements. Common recognition methods are as follows: Decision Tree (DT), Artificial Neural Networks (ANNs), Linear Discriminant Machine (LDM), and Support Vector Machine (SVM) (14, 15). An interpretable machine learning algorithm was used to evaluate the interrater reliability (IRR) of sleep stage annotation among sleep centers by Liu et al. (16). Intracenter and intercenter assessments were conducted on 679 patients without sleep apnea from six sleep centers in China's Taiwan. Bandyopadhyay and Goldstein (17) provided a concise overview of relevant terminology, definitions, and use cases of AI in sleep medicine. They conclude that artificial intelligence is a powerful tool in healthcare that may improve patient care, enhance diagnostic abilities, and augment the management of sleep disorders. Bozkurt et al. (18) proposed a new approach to sleep staging, which is one of the diagnostic procedures for the disease. An artificial intelligence-based sleep/awake system detection was developed for sleep staging processing. Photoplethysmography (PPG) signal and heart rate variable (HRV) were used in the study. Zhu et al. (19) adapted edge AI to develop a lightweight automatic sleep staging method for children using single-channel EEG. The trained sleep staging model will be deployed to edge smart devices so that the sleep staging can be implemented on edge devices which will greatly save network resources and improve the performance and privacy of the sleep staging application. Cesari et al. (1) conducted this study to evaluate interrater reliability between manual sleep stage scoring performed in two European sleep centers and automatic sleep stage scoring performed by the previously validated artificial intelligence-based Stanford-STAGES algorithm. Abbasi et al. (20) proposed an internet of things and ensemble-based automatic sleep stage classification in this study. Twelve EEG features, from nine bipolar channels, were used to train and test the base classifiers, including a convolutional neural network, support vector machine, and multilayer perceptron. Eldele et al. (21) proposed a novel attention-based deep learning architecture called AttnSleep to classify sleep stages using single-channel EEG signals. This architecture started with the feature extraction module based on a multi-resolution convolutional neural network (MRCNN) and adaptive feature recalibration (AFR). Prochazka et al. (22) devoted to the analysis of multi-channel biomedical signals acquired in the sleep laboratory. The goal of the study was to visualize the features associated with sleep stages as specified by an experienced neurologist in their adaptive classification. Delimayanti et al. (23) demonstrated the utilization of features extracted from EEG signals *via* FFT to improve the performance of automated sleep stage classification through machine learning methods. Gao et al. (24) developed a novel program called GI-SleepNet, a generative adversarial network (GAN)-assisted image-based sleep staging for mice that is accurate, versatile, compact, and easy to use. Loh et al. (25) proposed a deep learning model based on a one-dimensional convolutional neural network (1D-CNN) for CAP detection and homogenous 3-class classification for sleep stages: wakefulness (W), rapid eye movement (REM), and NREM sleep. The proposed model was developed using standardized EEG recordings. Manzano et al. (26) presented the first approximation using a single channel and information from the current epoch to perform the classification. The complete Physionet database has been used in the experiments. The main research objective of this paper is to realize the research on EEG sleep automatic staging by using signal processing method and deep learning theory, and establish an automatic sleep staging system with high accuracy and fast calculation speed to provide a more effective and convenient sleep staging methods for medical researchers, reduce the workload of clinical diagnosis, and reduce the error rate.

EEG signal

The brain is mainly responsible for controlling and regulating the thinking and conscious activities of the human body. EEG is a random signal reflecting the electrical activities of brain tissue. The cerebral cortex of the brain is the highest

center of the nervous system regulating body movement, which is responsible for the cognitive and emotional functions of the brain. Various human activities and senses can find corresponding regions in the cerebral cortex, and each region has its specific role and function. Neuron cells in the cerebral cortex are the reason why human beings have complex activities as advanced organisms. The condition for complex activities is that each neuron cell can communicate through dendritic connections, which makes human activities have countless possibilities (27).

Scientific research shows that EEG signal is formed by a large number of neuronal cells. It is a non-linear mixture of multiple cells connected by non-linear coupling. From the perspective of chaos theory, the human brain is a complex non-linear dynamic system. EEG signal has the following characteristics: weak signal and strong noise; high complexity of signal components; prominent frequency characteristics; non-linearity; and non-stationarity.

Composition of brain waves

The root cause of EEG is the response of neurons in the cerebral cortex to electrical activity. The signals generated by the potential activity of nerve cells can be obtained by placing electrodes on the scalp. The amplitude, frequency, and phase of the waveform contained in the EEG have certain characteristics. The bandwidth of EEG is 0.5–100 Hz, and the 0.5–30 Hz portion of spontaneous EEG is considered in clinical medicine. According to the frequency characteristics, it can be divided into four basic rhythm waves and non-basic waves. The regular waves in the sleep process are rhythm waves, and the irregular waves are non-basic waves. There are four kinds of rhythm waves:

(1) δ wave: 0.5–4 Hz, amplitude 20–200 μV. δ wave is the wave with the lowest frequency in the brain, which is obvious in the forehead. It appears more often in the EEG of infants or people with intellectual disabilities, and it will not occur in normal people when they are awake; however, this wave can be produced when adults are in extreme fatigue, deep sleep, and deep anesthesia. When the cortex is inhibited, δ rhythmic waves appear.
(2) θ wave: 4–8 Hz, amplitude about 20–150 μV. θ wave is a low-frequency rhythmic wave in the human brain. It is the main EEG activity in teenagers aged 10–17 years. It is obvious in temporal and parietal areas. Similar to δ wave, it occurs only when the brain nerve is in a sleepy, depressed, or frustrated state, and it does not occur when normal people are awake, mtf(Modulation Transfer Function) wave, if someone is awake, the signal contains a lot of θ wave, should be paid attention to, its brain may have lesions, which will lead to disease.
(3) α wave: 8–13 Hz, amplitude $<$100 μV. α wave is the most obvious waveform with a high frequency in adult EEG. It has a stable existence in the state of closed and relaxed eyes, but it disappears immediately after being stimulated by the outside world. The α wave represents human logical reasoning and creative thinking, which are not available to other animals. The α wave is a basic waveform of health and an important rhythmic wave reflecting various indexes of the brain.
(4) β wave: 13–30 Hz amplitude of 5–20 V. The frequency is 13–30 Hz. β wave is in the range of 20–30 Hz β two waves. β wave is the fastest rhythmic wave in the brain, which is mainly concentrated in the frontal and central regions. β waves are the manifestation of the excited state of the cerebral cortex and can also be used as the characteristics of brain awakening. It can be detected when people are moving, thinking, reading, or watching movies.

There are three types of common non-fundamental waves as follows:

(1) K-complex wave: there is no exact frequency range, and is composed of a positive and negative sharp waves. It first appears negative and then positive, and the transient time is $>$0.5 s. The K-complex wave usually appears together with the spindle wave, which are fused together to determine the S2 phase. The K-complex wave is obvious in the top and central regions.
(2) Sawtooth wave: the frequency range is 2–7 Hz. The waveform is like a sawtooth shape. Its existence generally represents the beginning of a dream, that is, sawtooth waves generally exist when the human body is in a light sleep state or dreaming.
(3) Spindle wave: the frequency is 12–14 Hz, and the amplitude is the largest and constantly changing in EEG. At first, it is the lowest, the highest in the center, and then gradually decreases to the lowest. It is obvious in the parietal and midline regions of the brain. The spindle wave is the main characteristic waveform during light sleep.

EEG detection

The EEG signal detection and recording equipment are composed of amplifier, electrode, display, etc. Electrodes are used to measure brain epidermal potential. Special electrodes need to be selected to measure EEG. In fact, the electrodes that can be used to measure brain potential include a tubular electrode, deep electrodes, disk electrodes, and so on. Different electrodes are designed to meet the different noise signals generated in the acquisition process of EEG signals. The conditions for selecting electrodes are as follows: small drift can reduce the noise interference of EEG; the polarization potential

is stable, indicating good reproducibility; small photoelectric effect ensures that the electrode is not affected by light and is not easy to aging; and good mechanical properties mean long electrode life. Considering the above factors, the disk electrode is the most commonly used electrode for EEG measurement.

Sleep stage

Sleep staging criteria

The EEG changes during sleep. According to different eye, EEG, and EMG signals, a sleep grading system can be designed to respond to different sleep levels. The R&K staging standard proposed in the 1960s is widely used in the world. Sleep is divided into six periods: wake, non-rapid eye movement (NREM, including I, II, III, and IV), and rapid eye movement (REM). This standard is of great significance to study sleep staging, which was once called the "gold standard." Later, researchers found that the characteristics of sleep stages III and IV were not obvious. Therefore, in 2007, the American Society of Sleep Medicine merged the two stages III and IV in the R&K standard, divided the total sleep period into five stages, and set the latest sleep staging standard (AASM standard), which is of epoch-making significance. The sleep staging standard selected in this paper is the AASM standard. Figure 1 presents the sleep staging criteria according to AASM.

Specific introduction to sleep stages

According to the AASM standard, the sleep process is divided into five stages: waking stage (W), rapid eye movement stage (REM), light sleep phase I (N1), light sleep phase II (N2), and deep sleep phase (N3). Each sleep stage will be briefly discussed below.

Waking stage (W period): adults are generally awake for about two-thirds of the time in a day. At this stage, eye movement is obvious, which is the characteristic of frequent blinking. During the awake period, the EEG shows an uninterrupted alpha wave (frequency 8–13 Hz, amplitude 20–100 μV).

Light sleep phase I (N1): this phase often occurs when the awake state changes to other sleep states, or when there is more body movement during sleep, generally accounting for about 5–10% of the whole night's sleep. The duration of each occurrence is short, often only 1–7 min. The characteristics of this period are as follows: the content of the alpha wave decreases (<50%), and the content of the theta wave (frequency is 4–8 Hz and the amplitude is 20–150 μV) will slowly increase.

Light sleep phase II (N2): this phase occupies a longer time than light sleep phase I during the whole night, generally accounting for 44–55% of the sleep duration. This period is characterized by the gradual emergence of two new characteristics of EEG signals: sleep spindle wave and K-complex wave. These waveforms are most obvious in the EEG-type detection of the central electrode. At this stage, the θ waves of EEG signals increase and gradually appear at the same time as the δ wave (frequency is 0.15–4 Hz and amplitude is 20–200 μV).

Deep sleep period (N3): when people enter deep sleep, that is, when the sleep stage is in N3, the repair functions of the body and the visceral function are the strongest. At this stage, the cerebrospinal fluid in the brain will regularly clean the "biological waste" in the brain, which is very important for stabilizing mood, restoring energy, and adjusting mental state. This stage accounts for about 10–20% of the total sleep time. It is the deepest stage in the sleep process. The approximate frequency range of EEG signals at this stage is 0.5–2 Hz, and δ waves slowly increased, accounting for more than 20% of the waves.

Rapid eye movement (REM): this accounts for 20–50% of the total sleep process. The biggest feature of the REM phase is rapid eye movement, which is similar to the EEG signal of the N1 phase, and the REM sawtooth phase occurs. At this stage, α waves increased but showed an irregular pattern, and were accompanied by rapid eye movement.

CNN-LSTM sleep stages algorithms

This paper will construct the automatic sleep stages model based on CNN and LSTM algorithms, train the model, and evaluate its classification performance through experiments.

Deep neural networks

Convolutional neural network (CNN)

A convolutional neural network (CNN) is a feedforward neural network, and it is featured with weight sharing and local reception. A typical CNN structure is shown in Figure 2. The convolution neural network is constructed by imitating the visual perception mechanism of biology, which can carry out supervised learning and unsupervised learning. The sharing of convolution kernel parameters in the hidden layer and the sparsity of inter-layer connections enable the convolution neural network to learn grid-like topology features with less computation. The function of the convolution layer is to extract the features of the input data. It contains multiple convolution kernels. Each element of the convolution kernel corresponds to a weight coefficient and a bias vector, which is similar to a neuron of a feedforward neural network. Each neuron in the convolution layer is connected to multiple neurons in the area

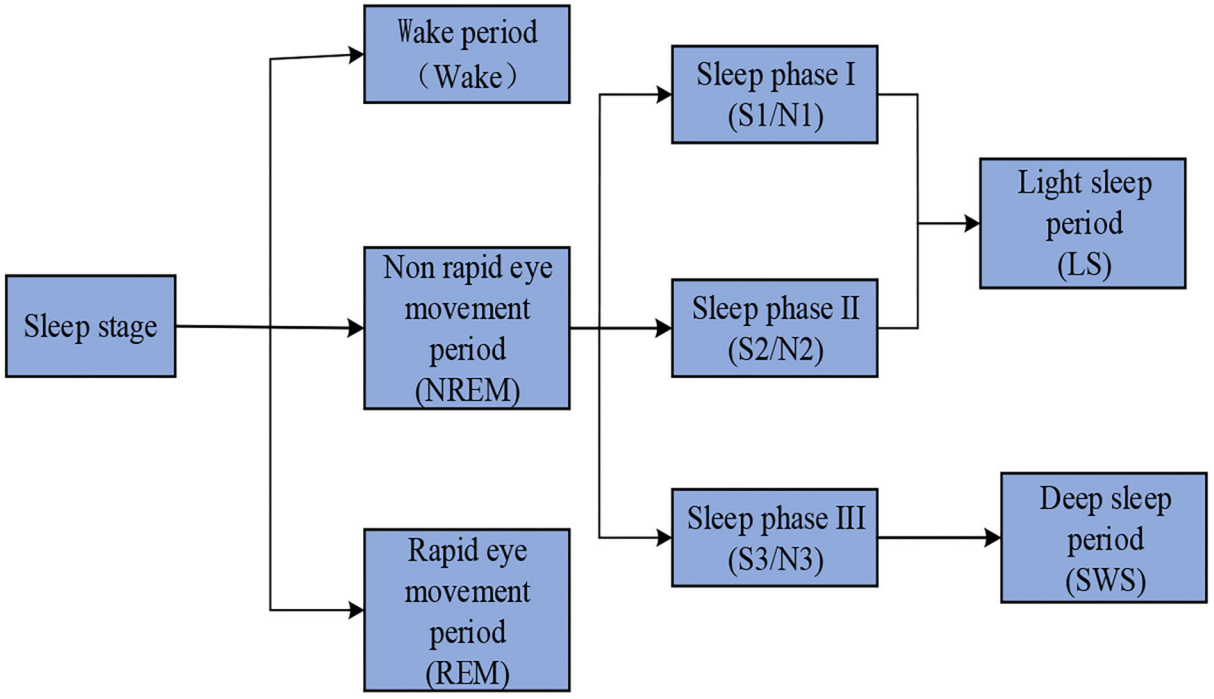


FIGURE 1
AASM sleep staging criteria.

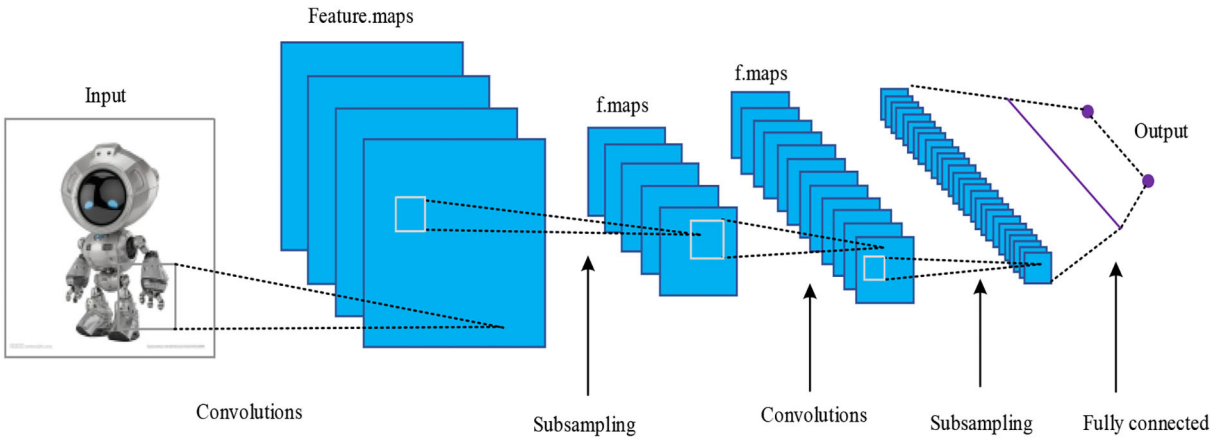


FIGURE 2
Typical CNN structure.

close to the previous layer, and the size of the area depends on the size of the convolution kernel.

In the structure, the output of the convolutional layer is locally connected to the input of the last layer, and it can be acquired by sliding the window gradually according to the step length. The corresponding local features with each parameter of the convolution kernel are multiplied and the weight of the kernel is kept unchanged, that is, weight sharing. Usually, in every convolutional layer, there are several convolution kernels and different features can be extracted. Through convolution, image features can be obtained, and after that, the classifiers can be trained according to these features directly, but such computation is huge and over-fitting occurs easily. Therefore, the addition of a pooling layer after the convolutional layer can shrink the feature map matrix and reduce the computation amount while enhancing the anti-jamming capability for feature

identification, e.g., image deformation and distortion. Through local connection, weight sharing, and pooling, CNN greatly reduces the complexity of the network.

The convolutional neural network is frequently used in the processing of images and discrete signals, and its network structure mainly includes the input layer, the convolutional layer, the pooling layer, the full-connected layer, and the output layer.

(1) Convolutional layer

The convolutional layer is the most important layer in CNN, and it extracts efficient high-level feature information from the data of the input layer and inputs the regions divided by sliding windows, and conducts convolution operations with the kernel parameters. During each computation, the kernel window translates at the fixed step length in the input layer until the entire input participates in the convolution operation. Then with the activation function, add non-linear factors and construct the feature map. The output of the computation is also called feature mapping. Assume the input to be a two-dimensional tensor I and a two-dimensional convolution kernel K, CNN uses the cross-correlation function which is basically the same as the convolution operation and does not reserve the convolution kernel.

$$\begin{aligned} S(i,j) &= (I * K)(i,j) \\ &= \sum_m \sum_n I(m,n) K(i-m, j-n) \\ &= \sum_m \sum_n I(i+m, j+n) K(m,n) \end{aligned} \tag{1}$$

In essence, CNN's convolution operation gets the local features from the weighted sum of local input through the convolution kernel and its convolution operation is as follows:

Convolution operation has strengths, such as sparse connection and parameter sharing. In the networks which connect all layers with a full-connected layer, each input and output cells are combined and connected, resulting in huge computation and easy over-fitting, exploding and vanishing gradient, and other issues. However, both the size and quantity of the convolution kernels are much fewer than the input nerve cell, and they can significantly reduce the number of weight parameters and have the characteristics of sparse weight. Besides, every weight factor in the convolution kernel has translated convolution operation with all neurons of the last layer, reducing the dense matrix operation in the full-connected layer and increasing the efficiency of reverse update of network parameters through weight and parameter sharing.

(2) Pooling layer

The pooling layer processes the output of the last layer through pooling functions, such as the maximum, mean, and L^2 norm. It comprehensively takes into account the neighbor value of each position in the output of the last layer, conducts further feature localization processing on the original feature map, and maps it into another feature space. The pooling layer can reduce the scale of input, maintain the output of the features in the last layer, and effectively lower the dimensions of features so as to enhance the computational efficiency of the network.

(3) Full-connected layer

Full-connected layer maps the feature map output from previous layers into a feature vector and reduces the impact of spatial location on object identification. As the full-connected layer is based on dense matrix operation, it has low computational efficiency, and it is easy to contribute to over-fitting; therefore, Dropout is often used to increase the sparsity of the full-connected layer and reduce the feature redundancy among neurons.

LSTM network

A recurrent neural network (RNN) is a neural network that is suitable to be used in the processing of sequence data. RNN increases the lateral ties among cells, making it preserve the information state of the context windows when analyzing time sequence data, and it has been widely applied in natural language processing and machine vision which includes serial input. In practical applications, it mainly includes long short-term memory (LSTM) and gated recurrent cell (GRU).

This paper mainly uses LSTM to acquire the time sequence features of sleep EEG. LSTM introduces logic control gates, and they are responsible for the memory and update of the history state. The input gate is responsible to control the impact of the input x_t on the current LSTM cell state c_t, and the forgotten gate controls the impact of the state c_{t-1} of the last moment on c_t. The cell structure of LSTM is shown in Figure 3.

The expressions among the LSTM forgotten door f_t, input door i_t,output door o_t, and state vector $\tilde{c}_t$ within the cell can be given by:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

$$i_t = \sigma\left(W_t \cdot \left[h_{t-1}, x_t\right] + b_t\right) \tag{3}$$

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{4}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \tag{5}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{6}$$

$$h_t = o_t \times \tanh(c_t) \tag{7}$$

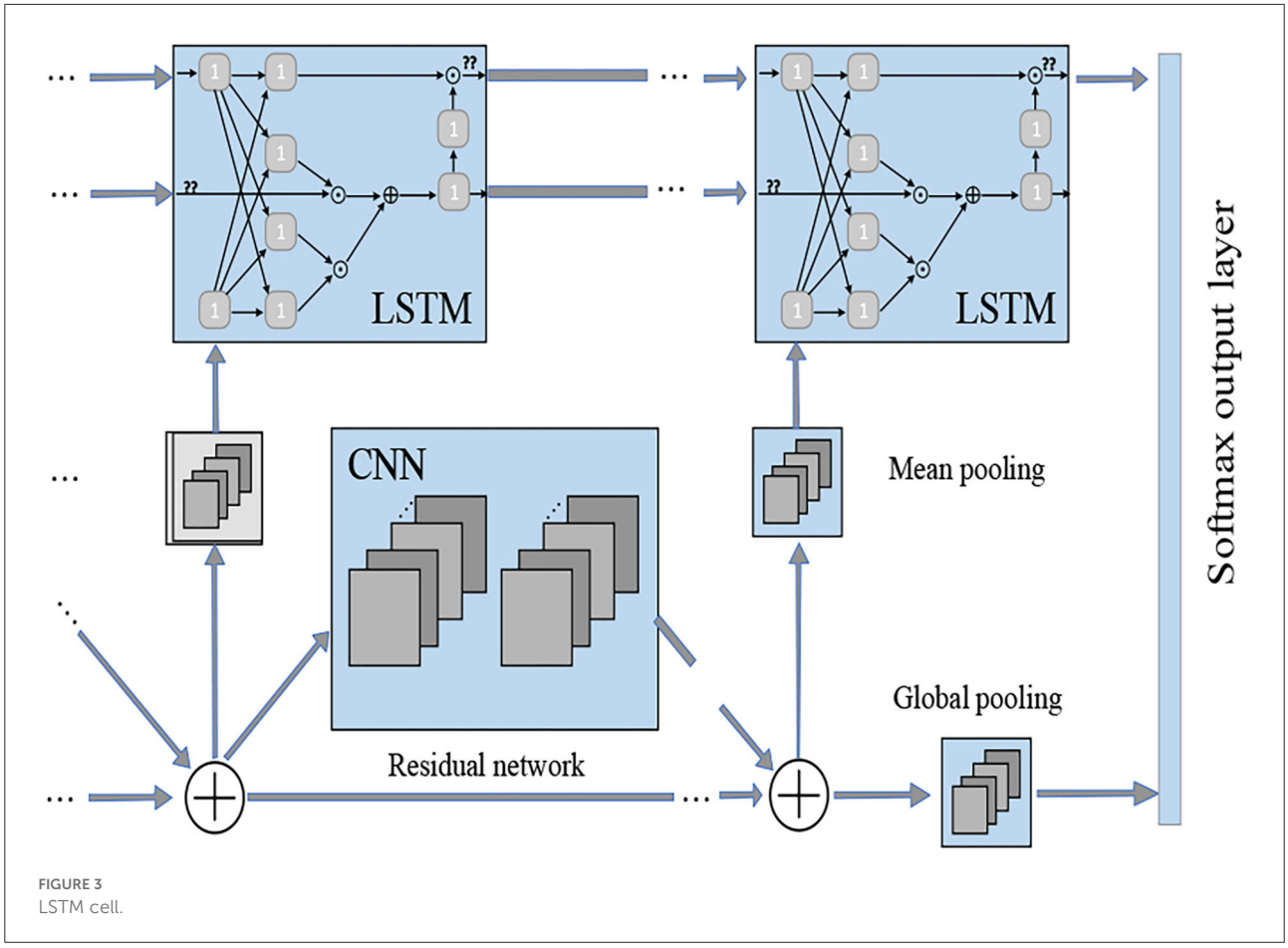


FIGURE 3
LSTM cell.

where $W_\bullet$ is the weight matrix, $b_\bullet$ is the bias vector, σ is logistic sigmoid function, and tanh is the hyperbolic tangent mapping activation function.

Building of automatic sleep stages model based on neural network

CNN-LSTM automatic sleep stages model

The identification CNN-LSTM activity can be done by introducing the LSTM model on the basis of the CNN model, and it uses CNN to process feature extraction and LSTM to process long-time sequence data. CNN has a two-layered structure: one is the feature extraction layer and its structure adopts the ReLU function as the activation function of the convolutional network, and the other is the feature mapping layer and its structure uses the sigmoid function as the activation function. It takes the pre-processed activity sequence as the input of the model and sets the convolution kernel with a length of l. After convolution operation, it gets the new feature sequence with a length of $le = (n - l + 1)$, which is shortened through the maximum pooling process, and finally obtains the extracted activity feature vector $v = (v_1, v_2, \ldots, v_g)$. As the extension of the recurrent neural network, LSTM can better solve the strong data dependency problem in time sequence. The specific model of the LSTM model in activity identification is indicated in Figure 4.

Figure 4 shows the structure of how the single LSTM neuron processes data dependencies. Therefore, LSTM replaces the full-connected layer of CNN and takes the activity feature vector $x = \{v_1, v_2, \ldots, v_g\} = \{x_1, x_2, \ldots, x_g\}$ obtained as the LSTM input with each feature vector corresponding to an activity label A_i. LSTM determines the reservation of cell state according to the control among its input gate, forgotten gate, and output gate, maintains the relationship between the data with strong dependencies, and solves long-term dependency problems in order to accurately identify the activities. $c(t-1)$ represents the unit state of the last moment, and the philosophy of LSTM is that the state of the current moment t depends on the state of the same neuron in the last moment $t-1$. $h(t)$ can be defined as the output state of the hidden layer at the current moment, and it can be given by

$$h(t) = \log h(t-1) \bullet x(t) \tag{8}$$

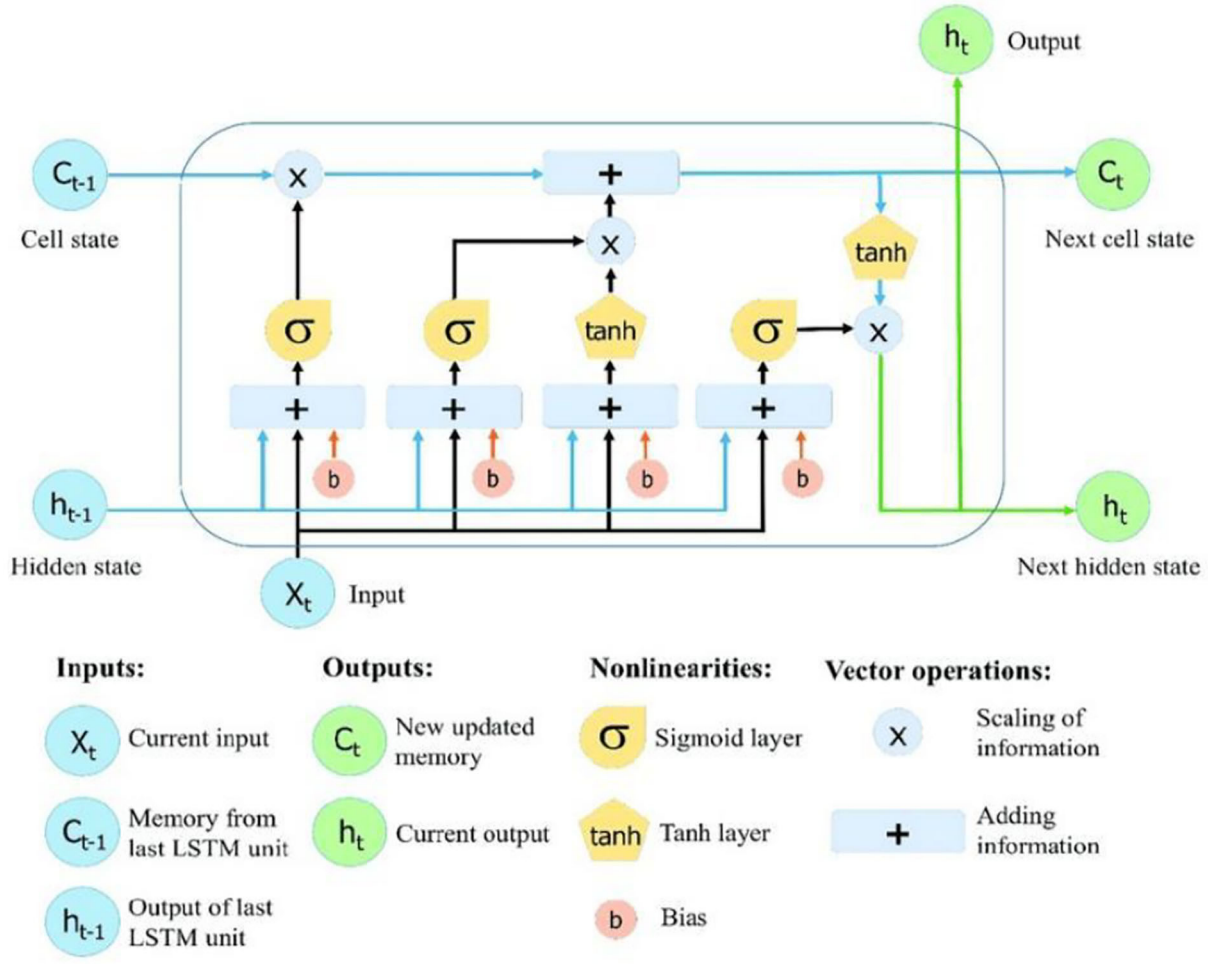


FIGURE 4
Single LSTM neuron processing data dependencies.

where $x(t)$ represents the input data of neuron at moment t and $h(t-1)$ means the LSTM output of the last moment, i.e.,.. The values after forgotten threshold, input threshold, and output threshold can be calculated by:

$$f(t) = \sigma\left(W_{(f)} \bullet x(t) + U_{(f)} \bullet h(t-1) + b_{(f)}\right) \quad (9)$$

$$i(t) = \sigma\left(W_{(i)} \bullet x(t) + U_{(i)} \bullet h(t-1) + b_{(i)}\right) \quad (10)$$

$$o(t) = \sigma\left(W_{(o)} \bullet x(t) + U_{(o)} \bullet h(t-1) + b_{(o)}\right) \quad (11)$$

where $f(t)$, $i(t)$, and $o(t)$ represent the output state of the corresponding thresholds, respectively. W and U are the weights of each branch, and b is the bias of each threshold. The activation sigmoid function of input gate decides what information to update, and tanh layer creates a new cell state, which can be given by:

$$\tilde{c}(t) = \tanh(W(c) \bullet x(t) + U(c) \bullet h(t-1) + b(c)) \quad (12)$$

Meanwhile, the old cell state will be updated to the new cell state $c(t)$, which can be given by:

$$c(t) = f(t) \bullet c(t-1) + i(t) \bullet \tilde{c}(t) \quad (13)$$

The sigmoid function of the output gate determines what information to output and obtains the output $h(t)$ of the current hidden layer through the tanh function, as suggested by Formula (14):

$$h(t) = o(t) \bullet \tanh\left(c(t-1)\right) \quad (14)$$

In the entire LSTM computation process, the output $h(t-1)$ of the hidden layer of the neuron at moment $t-1$ and $x(t)$ are the input of the neuron at moment t. The input of the neuron in the input layer produces the neuron output $h(t)$ under the action of activation function, and the memory state $c(t-1)$ of the neuron cell at moment $t-1$ forms that memory state $c(t)$ at moment t. In the learning process, Formulas (5) and (7) show that the activation function tanh affects $c(t)$, and the affected $c(t)$ has point multiplication with the output state $o(t)$ after the control of the output threshold to get the output $h(t)$ of the hidden layer at moment t. The Softmax activation function normalizes the value of the weight α of the hidden layer, gets the multiplication product and sum of the weight $\alpha(i)$ in the hidden layer with the output $h(t)$ of the hidden layer at the current moment, and obtains the final feature expression $y(t)$ of the activity identification samples so as to get the respective probabilities of the activities in the sample set, as shown in Formula (15).

$$y(t) = \sum \alpha(i)h(t) \tag{15}$$

Therefore, the interaction among the LSTM input gate, output gate, and forgotten gate decides whether the information should be preserved or abandoned in order to solve the problem of strong data dependencies. This cycle is repeated and after the iteration of one batch (one-time training samples), the accuracy is calculated as the difference between the expected and actual activities.

Model structure of automatic sleep stage algorithm

This paper builds the automatic sleep stage algorithm model, mainly including the model structure, size of convolution kernel, and so on. Sleep EEG screens and produces IMF through EMD, and it can effectively present the rhythm wave form and frequency feature of EEG. Therefore, the CNN structure is used to get the time-frequency domain features of each group of IMF, and these features will help with the decision-making of the model on sleep stages and are used to improve the performance of the CNN-LSTM model.

To better simulate the stages of sleep, CNN input includes the EEG and EOG signals which are not processed in the classification stage, and takes a digital matrix as the input. The proposed model in this paper adopts four two-dimensional convolution layers and uses rectified linear unit (ReLU) as the activation function. For each two-dimensional convolution layer, it uses dropout to avoid over-fitting, resulting in neurons being closed or abandoned at a 20% probability. Every convolution layer is followed by one maximum pooling layer before two full-connected layers. The first connected layer has 4,096 cells, and the second one has 1,500 cells. Both layers also adopt dropout to avoid over-fitting and lead to closed or abandoned neurons at a 50% probability. The last is a full-connected layer with a size of 5, and it chooses ReLU as its activation function. All activation functions used in the convolution layers are ReLU with a negative slope of 0.

In EEG-CNN automatic sleep stage model, the input is a 30-s EEG signal with a sampling rate of 200 Hz, and it is represented in a 5 * 6,000 matrix form. Here 5 means that it adopts the data of five channels, and each EEG sample signal is built alternatively in the order of input layer, Cov, Dropout, Pool until FC1 and FC2, and output. This paper divides EEG and EOG signals of the sleep stages of all types into a test set and training set. To be specific, 20% of all datasets are randomly selected as the test set, and the remaining are selected as the training set in each training stage. When using CNN at a fixed-length time sequence, the output of the convolution directly correlates with the input, the number of convolution layers, and their step lengths. If the output of the last convolution layer is too big, then the majority of the weights will be located in the full-connected layer. CNN connects with LSTM in a parallel manner, and it is used to learn the time sequence features of the output features of two networks. The network output is five sleep stages divided by the AASM sleep staging rule. Figure 5 represents the neural network model of sleep stages.

In Figure 5, the output dimension and core size of each layer are as follows: input layer: 300 × 2; 1st LSTM layer: 300 × 64; 2nd LSTM layer: 64; 1st convolution layer: 296 × 64,5; 1st pooling layer: 98 × 64,3; 2nd convolution layer: 94 × 32,5; 2nd pooling layer: 31 × 32,3; flatten: 992; concatenate: 1,056; 1st full-connected layer: 64; 2nd full-connected layer: 32; and Softmax: 5.

Every convolution layer in this model includes three operations: convolution operation, batch normalization, and activation function non-linearization. Batch normalization acts behind convolution operation, and it is used to re-adjust the distribution of the output in the last layer. It standardizes the input of each hidden layer and prevents the gradient diffusion caused by network deepening. The input layer of the CNN-LSTM designed in this study has 300 * 2 dimensions. It extracts time sequence features through LSTM, captures profound local correlation information through parallel CNN structure, fuses the convolution features and time sequence features into the full-connected layer, and outputs the multi-classification result from the Softmax layer. The size of batch processing used in this experiment is set to 20. It selects ReLU as the activation function, which is a piecewise linear function and has unilateral inhibition, making the neurons have sparse activation and play the part of automatic decoupling, and increase the convergence speed. The function $f(x)$ can be given by:

$$f(x) = \max(0, x) \tag{16}$$

Then, the full-connected layer maps the high-dimensionality and high-level features obtained from the previous model into

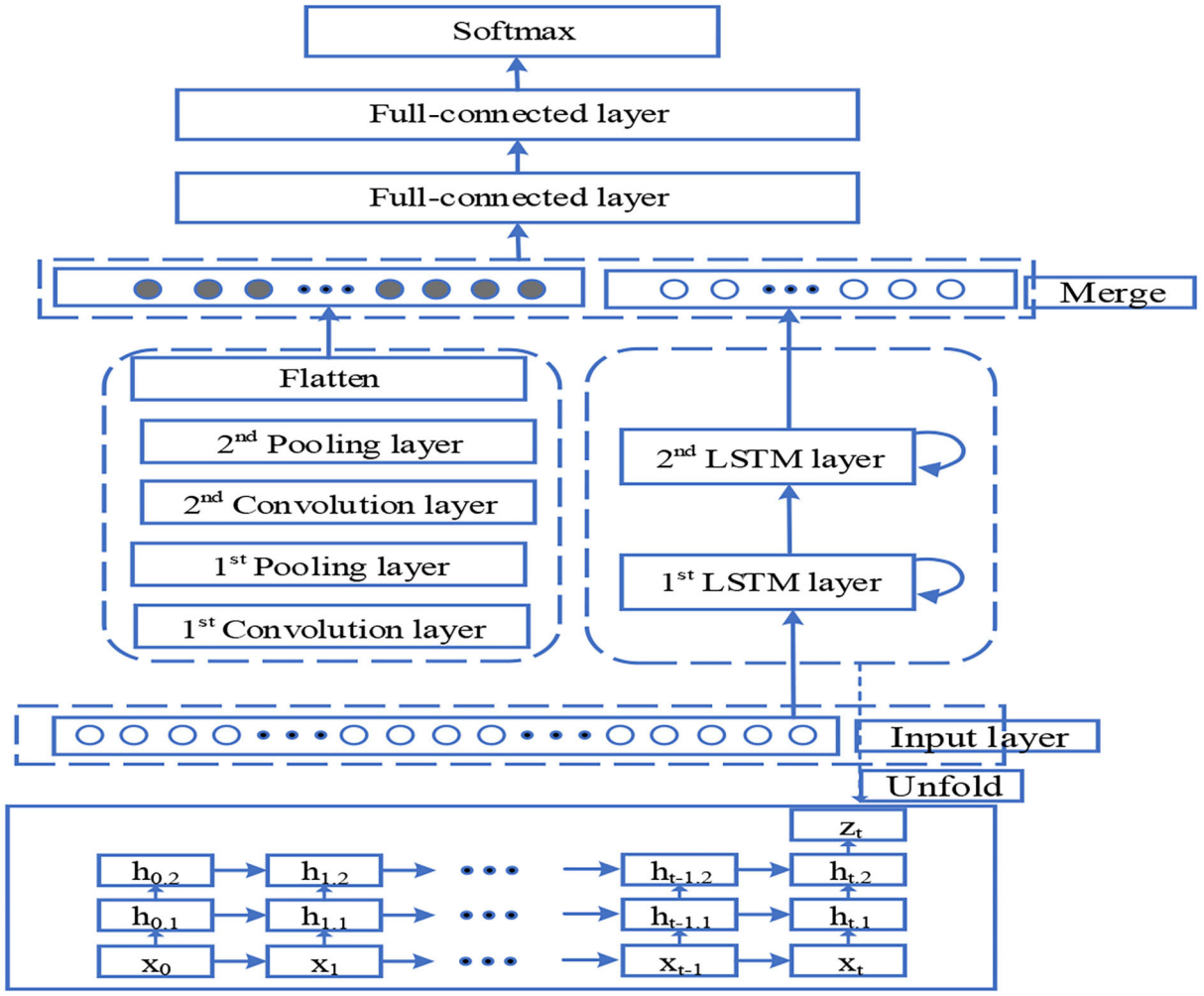


FIGURE 5
Neural network model of sleep stages.

the label space with five sleep stages. This model finally uses the Softmax classifier, into which the full-connected layer inputs the five real numbers obtained by multiplying the input feature vector and weight matrix and adding the bias input. The Softmax layer converts it into the probability of the corresponding five sleep stages, and the model judges the staging result according to the maximum probability.

Contrast experiment and result analysis

Performance evaluation index

Recall (REC), Precision (PRE), F1 (F1-Score, F1), and ACC (Accuracy) are employed to evaluate the classification performance of the proposed model. ACC is defined as the proportion of the number of samples of the prediction pair to the total number of samples. PRE indicates how many of the predicted positive samples are really positive samples. REC indicates how many positive examples in the sample are correctly predicted. They are common indicators in the statistical analysis. MF1 refers to the mean value of all F1 values in each sleep stage. ACC stands for the percentage of the classification of accurate sleep stages to all sleep stages. REC, PRE, F1, ACC, and MF1 are calculated with the following formulas respectively:

$$PRE = \frac{TP}{TP + FP}, REC = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = 2\frac{PRE * REC}{PRE + REC} \quad (18)$$

TABLE 1 The subject division of training set and test set.

Data division	Training set	Test set
Subject Reference No.	7–30	1–6

TABLE 2 Number of 30-s epochs for each sleep stage in training set.

Sleep stages	W	N1	N2	N3	REM
Number	9,563	3,177	17,246	4,854	7,393

TABLE 3 Number of 30-s epochs for each sleep stage in test set.

Sleep stages	W	N1	N2	N3	REM
Number	2,324	1,069	4,835	1,378	1,941

$$ACC = \frac{\Sigma_{C=1}^{C} TP_C}{N} \tag{19}$$

$$MF_1 = \frac{\Sigma_{C=1}^{C} F1_C}{C} \tag{20}$$

where TP_c is the number with accurate prediction in all classes and N is the total number of samples in all 30 s. $F1_c$ is the score of $F1$ of a certain class, and C is the number of sleep stages. TP is true positive, FN is false negative, and FP is false positive.

Training process

The experiment uses the data of 30 subjects in the MIT-BIH dataset, totaling 60 all-night sleep data. In the experiment, the use of LSTM enables the extraction of the features in a time sequence. In order not to interrupt the time sequence of the sleep stages, a division is made to the data with each subject as the unit. Table 1 presents the division of the subjects in the training set and test set.

The training set is made of the sleep data of 24 people for two nights, totaling 48 nights. The number of samples in each sleep stage in the training set is listed in Table 2.

The test set comprises the sleep data of two nights of six people, totaling 12 nights. The samples of each sleep stage are presented in Table 3.

In the training process, evaluate and validate the automatic sleep stage model apart from establishing the test set for the data of six persons, which is totally independent of the training set. Divide the sleep data of 24 persons in the training set into 12 on average, each of which includes the sleep data of four nights. In the 10th round of training, extract and use a ground subset as the validation set.

In the pre-training phase, randomly resample the data in W, N1, N3, and REM stages and perform under-sampled processing in the N2 stage, making the number of samples from six sleep stages every night not to exceed 200 after resampling and under-sampling. The total generations of data training have 100 epochs. A stochastic gradient descent optimizer is used and the initial learning rate is 0.015, which decays 0.1-fold after iterations and training of 20 generations. Set the batch as 10.

Experiment result and analysis

Hilbert transform focuses on extracting the local characteristics of the signal. HHT amplitude spectrum and marginal spectrum can obtain the time-frequency distribution characteristics of each IMF and the global energy distribution of each frequency. They are distributed in a narrow-band frequency, and the central frequency has a certain distance. It can be seen that the energy of IMF in a local frequency band is relatively concentrated, and the energy proportion of the high-frequency component is relatively small. The multi-resolution characteristics of HHT make the frequency and energy characteristics of this part of the signal more obvious. In order to study the energy distribution characteristics of each sleep period, the data of subjects for 10 days are randomly selected from the MIT-BIH dataset. Forty-five EEG signals are taken from each sleep period every day, and a total of 450 EEG signals are taken during all sleep periods. The instantaneous frequencies of the first four IMF components after empirical mode analysis are calculated, the frequency points are counted, and the probability density histogram is drawn, as shown in Figure 6.

This experiment adopts the EGG in MIT-BIH dataset as the input research subject of the model and performs automatic sleep stage model training on the training method and the divided dataset. The goal of the model is to classify sleep into five stages: W, N1, N2, N3, and REM according to AASM rules. In the training steps, the loss and accuracy curves of the training set and test set are shown in Figures 7, 8.

As shown in the performance curves of the training process, the loss function can converge quickly, and the total accuracy reaches 90.6% when the data finishes 100 iterations. But for the test set, its loss function value vibrates greatly. The study shows that with multi-channel EEG and EOG, it is viable to classify the sleep stages with the CNN-LSTM in this study and that EOG data should not be ignored in the study of automatic staging. The training is end-to-end and needs no professional knowledge to select any signal preprocessing method, which is a strength. Neural networks can independently learn the features of each stage and is the most suitable method for classification tasks. From the features of convolutional networks, the convolution layer can learn the excellent filter. Table 4 presents the comparison of performance between the proposed model and other models.

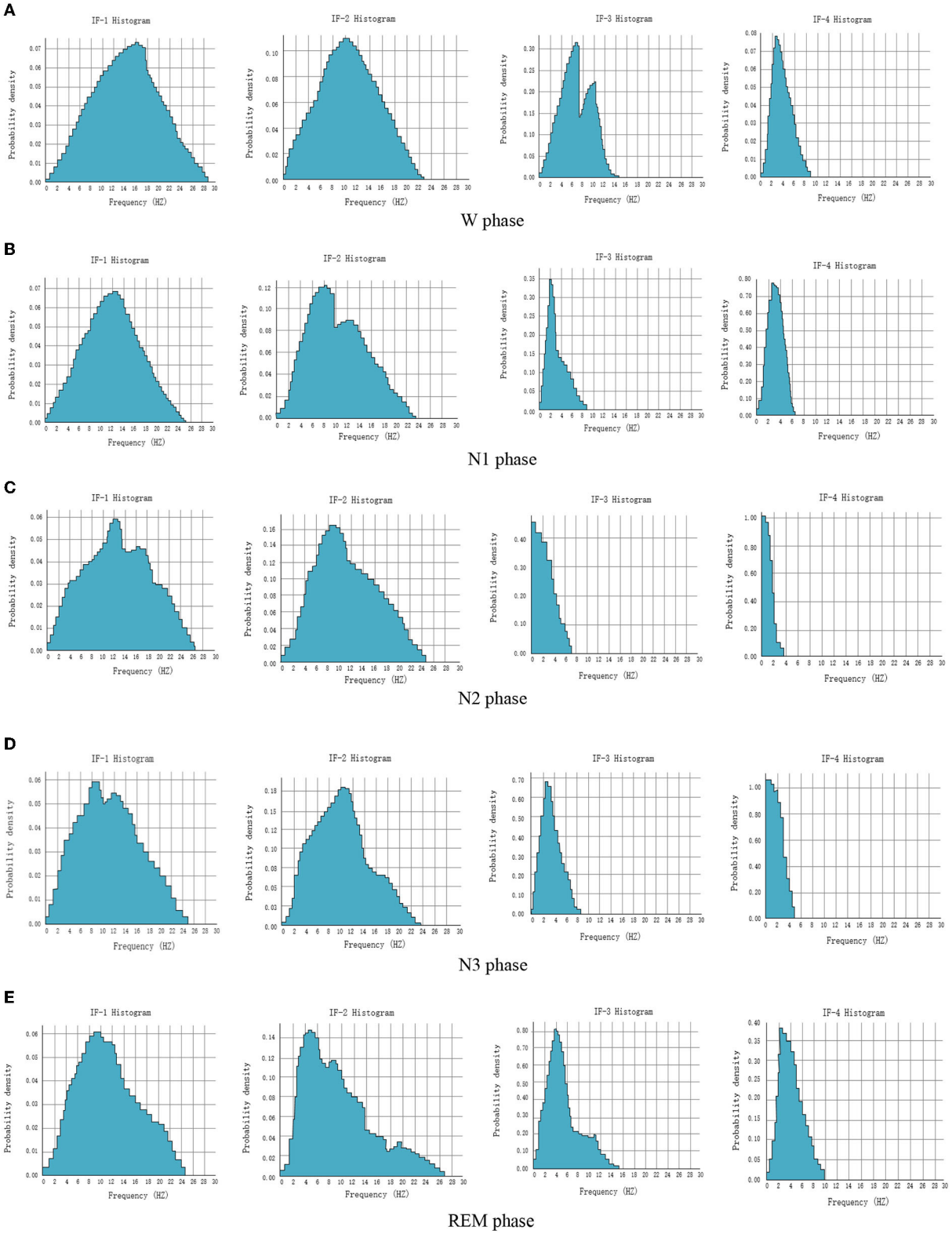


FIGURE 6
Probability density histogram of instantaneous frequency of EEG signals in each sleep period. **(A)** W phase. **(B)** N1 phase. **(C)** N2 phase. **(D)** N3 phase. **(E)** REM phase.

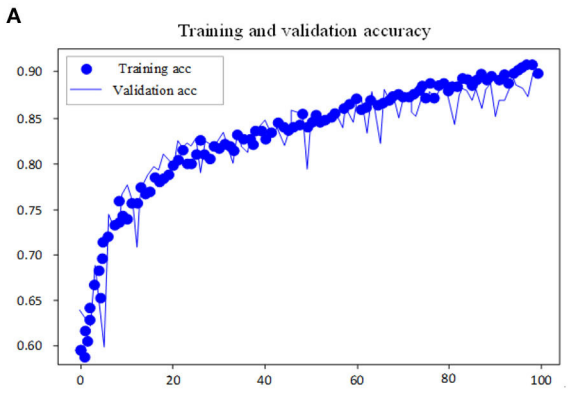


FIGURE 7
Accuracy and loss curve of pre-training process. **(A)** Training and validation accuracy. **(B)** Training and validation loss.

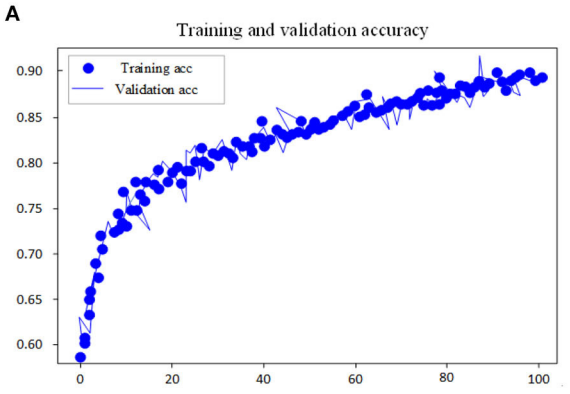


FIGURE 8
Accuracy and loss curve of fine tuning process. **(A)** Training and validation accuracy. **(B)** Training and validation loss.

TABLE 4 Performance comparison.

Author	Method	F1 scores (%)					Accuracy (%)
		W	N1	N2	N3	REM	
Fraiwan et al. (2015) (28)	Time-frequency analysis of and random forest classifier	88.32	53.26	85.28	86.31	85.15	84.72
Tsinalis et al. (2016) (29)	Time-frequency analysis and stacked sparse autoencoders	83.47	48.65	85.09	84.28	83.24	81.40
Sors et al. (2018) (30)	Convolutional neural networks	96.86	35.42	89.41	85.52	82.56	89.33
Ours	Multi-channels EEG & EOG CNN-LSTM	90.73	59.78	90.55	83.22	87.38	90.60

It can be seen from Table 4 that the proposed model shows quite good performance and better staging accuracy than other models, which demonstrates the generalization ability of the proposed model. As shown in the results, the multi-channel EEG and EOG model, adopted in this study, is better than the existing multi-channel models. Although other models show excellent performance, the performance of the proposed model is better than other models. From the staging results of different models, it can be concluded that the proposed model has significant advantages.

Conclusion and future work

In this study, an intelligent automatic sleep staging model based on CNN and LSTM is proposed. As sleep is an extremely complicated process, it is difficult to obtain high accuracy using automatic staging methods (31, 32). The proposed automatic sleep staging method, which is based on EEG, still needs improvement, mainly in the following aspects: (1) The impact of other physiological parameters should be considered. As sleep is a complex process, physiological parameters, such as electrocardio, myoelectricity, electro-oculogram, and breath, have a certain influence on sleep (33). Hence, more parameters should be introduced into the study of sleep stages, which can further increase the accuracy of automatic sleep staging. (2) More feature parameters and experiment samples need to be used to extract more feature parameters (34), including time domain, frequency domain, and non-linear features. Multi-channel data should be selected and more features and data need to be used to make a comprehensive study of sleep stages.

In the future work, we will further explore the application of neural network technology to the automatic sleep staging task, improve the overall accuracy, and provide a theoretical basis for the application of various neural networks to sleep staging.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LZ and MD designed the experiments and research project. YZ performed the experiments and analyzed the data. LS wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the National Science Foundation of China (Grant No. 61772190).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Cesari M, Stefani A, Penzel T, Ibrahim A, Hackner H, Heidbreder A, et al. Inter rater sleep stage scoring reliability between manual scoring from two European sleep centers and automatic scoring performed by the artificial intelligence-based Stanford-STAGES algorithm. *J Clin Sleep Med*. (2021) 17:1237–47. doi: 10.5664/jcsm.9174

2. Chriskos P, Frantzidis CA, Gkivogkli PT, Bamidis PD, Kourtidou-Papadeli C. Automatic sleep staging employing convolutional neural networks and cortical connectivity images. *IEEE Trans Neural Netw Learn Syst*. (2020) 31:113–23. doi: 10.1109/TNNLS.2019.2899781

3. Zhu TQ, Luo W, Yu F. Multi-branch convolutional neural network for automatic sleep stage classification with embedded stage refinement and residual attention channel fusion. *Sensors*. (2020) 20:9384–99. doi: 10.3390/s20226592

4. M. F. Leung, J. Wang. Minimax and bi-objective portfolio selection based on collaborative neurodynamic optimization. *IEEE Trans Neural Netw Learn Syst*. (2021) 32:2825–36. doi: 10.1109/TNNLS.2019.2957105

5. Abdollahpour M, Rezaii TY, Farzamnia A, Saad I. Transfer learning convolutional neural network for sleep stage classification using two-stage data fusion framework. *IEEE Access*. (2020) 8:180618–32. doi: 10.1109/ACCESS.2020.3027289

6. Gao X, Li YR, Lin GD, Xu MK, Zhang XQ, Shi YH, et al. The accuracy and influencing factors of sleep staging based on single-channel EEG via a deep neural network. *Chin J Otorhinolaryngol Head Neck Surg*. (2021) 56:1256–62. doi: 10.3760/cma.j.cn115330-20210513-00267

7. Sholeyan AE, Rahatabad FN, Setarehdan SK. Sleep staging with deep structured neural net using gabor layer and data augmentation. *Turk J Electr Eng Comput Sci*. (2021) 29:2920–33. doi: 10.3906/elk-1908-50

8. Jia ZY, Lin YF, Wang J, Ning XJ, He YL, Zhou RH, et al. Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. *IEEE Trans Neural Syst Rehabil Eng*. (2021) 29:1977–86. doi: 10.1109/TNSRE.2021.3110665

9. Hong H, Chi ZR, Feng DD, Grunstein R, Gordon C. A residual based attention model for EEG based sleep staging. *IEEE J Biomed Health Inform*. (2020) 24:2833–43. doi: 10.1109/JBHI.2020.2978004

10. Cai Q, Gao ZK, An JP, Gao S, Grebogi C. A graph-temporal fused dual-input convolutional neural network for detecting sleep stages from EEG signals. *IEEE Trans Circuits Syst Ii-Express Briefs*. (2021) 68:777–81. doi: 10.1109/TCSII.2020.3014514

11. Ye JA, Xiao QF, Wang J, Zhang HJ, Deng JX, Lin YF. CoSleep: a multi-view representation learning framework for self-supervised learning of sleep stage classification. *IEEE Signal Process Lett*. (2022) 29:189–93. doi: 10.1109/LSP.2021.3130826

12. Sun CL, Chen C, Li W, Fan JH, Chen W. A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning. *IEEE J Biomed Health Inform*. (2020) 24:1351–66. doi: 10.1109/JBHI.2019.2937558

13. Sokolovsky M, Guerrero F, Paisarnsrisomsuk S, Ruiz C, Alvarez SA. Deep learning for automated feature discovery and classification of sleep stages. *IEEE-ACM Trans Comput Biol Bioinform*. (2020) 17:1835–45. doi: 10.1109/TCBB.2019.2912955

14. Phan H, Andreotti F, Cooray N, Chen OY, De Vos M. Joint classification and prediction CNN framework for automatic sleep stage classification. *IEEE Trans Biomed Eng*. (2019) 66:1285–96. doi: 10.1109/TBME.2018.2872652

15. Zhang JM, Wu Y. Automatic sleep stage classification of single-channel EEG by using complex-valued convolutional neural network. *Biomed Eng-Biomedizinische Technik.* (2018) 63:177–90. doi: 10.1515/bmt-2016-0156

16. Liu GR, Lin TY, Wu HT, Sheu YC, Liu CL, Liu WT, et al. Large-scale assessment of consistency in sleep stage scoring rules among multiple sleep centers using an interpretable machine learning algorithm. *J Clin Sleep Med.* (2021) 17:159–66. doi: 10.5664/jcsm.8820

17. Bandyopadhyay A, Goldstein C. Clinical applications of artificial intelligence in sleep medicine: a sleep clinician's perspective. *Sleep Breathing.* (2022) 8:211–28. doi: 10.1007/s11325-022-02592-4

18. Bozkurt MR, Ucar MK, Bozkurt F. Bilgin, C. Development of hybrid artificial intelligence based automatic sleep/awake detection. *IET Sci Measur Technol.* (2020) 14:353–66. doi: 10.1049/iet-smt.2019.0034

19. Zhu LQ, Wang CM, He ZH, Zhang Y, A. Lightweight automatic sleep staging method for children using single-channel EEG based on edge artificial intelligence. *World Wide Web-Internet Web Inform Syst.* (2021) 8:110–21. doi: 10.1007/s11280-021-00983-3

20. Abbasi SF, Jamil H, Chen W. EEG-based neonatal sleep stage classification using ensemble learning. *CMC-Comput Mater Continua.* (2020) 70:4619–33. doi: 10.32604/cmc.2022.020318

21. Eldele E, Chen ZH, Liu CY, Wu M, Kwoh CK Li XL, Guan CT. An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng.* (2021) 29:809–18. doi: 10.1109/TNSRE.2021.3076234

22. Prochazka A, Kuchynka J, Vysata O, Schatz M, Yadollahi M, Sanei S, et al. Sleep scoring using polysomnography data features. *Signal Image Video Process.* (2018) 12:1043–51. doi: 10.1007/s11760-018-1252-6

23. Delimayanti MK, Purnama B, Nguyen NG, Faisal MR, Mahmudah KR, Indriani F, et al. Classification of brainwaves for sleep stages by high-dimensional FFT features from EEG signals. *Appl Sci.* (2020) 10:217–20. doi: 10.3390/app10051797

24. Gao T, Li J, Watanabe Y, Hung C, Yamanaka A, Horie K, et al. GI-SleepNet: a highly versatile image-based sleep classification using a deep learning algorithm. *Clocks Sleep.* (2020) 3:581–97. doi: 10.3390/clockssleep3040041

25. Loh HW, Ooi CP, Dhok SG, Sharma M, Bhurane AA, Acharya UR. Automated detection of cyclic alternating pattern and classification of sleep stages using deep neural network. *Appl Intell.* (2022) 55:2903–17. doi: 10.1007/s10489-021-02597-8

26. Leung MF, Wang J. Cardinality-constrained portfolio selection based on collaborative neurodynamic optimization. *Neural Netw.* (2022) 145: 68–79. doi: 10.1016/j.neunet.2021.10.007

27. Shi M, Yang C, Zhang, D. A smart detection method of sleep quality using EEG signal and long short-term memory model. *Math Probl Eng.* (2021) 2021:1412–25. doi: 10.1155/2021/5515100

28. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhause, H. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput Methods Programs Biomed.* (2015) 108:10–9. doi: 10.1016/j.cmpb.2011.11.005

29. Tsinalis O, Matthews PM, Guo Y. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. *Ann Biomed Eng.* (2016) 44:1587–97. doi: 10.1007/s10439-015-1444-y

30. Sors A, Bonnet S. Mirek S, Vercueil L, Payen, J-F. A convolutional neural network for sleep stage scoring from raw single-channel EEG *Biomed Signal Process Control.* (2018) 42:107–14. doi: 10.1016/j.bspc.2017.12.001

31. Sun CL, Fan NH, Chen C, Li W, Chen W, A. Two-stage neural network for sleep stage classification based on feature learning, sequence learning and data augmentation. *IEEE Access.* (2019) 7:109386–97. doi: 10.1109/ACCESS.2019.2933814

32. Caballero HSG, Westenberg MA, Gebre B, van Wijk JJ. V-awake: a visual analytics approach for correcting sleep predictions from deep learning models. *Comput Graph Forum.* (2019) 38:1–12. doi: 10.1111/cgf.13667

33. Korkalainen H, Leppanen T, Duce B, Kainulainen S, Aakko J, Leino A, et al. Detailed assessment of sleep architecture with deep learning and shorter epoch-to-epoch duration reveals sleep fragmentation of patients with obstructive sleep apnea. *IEEE J Biomed Health Inform.* (2021) 25:2567–74. doi: 10.1109/JBHI.2020.3043507

34. Chriskos P, Kaitalidou DS, Karakasis G, Frantzidis C, Gkivogkli PT, Bamidis P, et al. Automatic sleep stage classification applying machine learning algorithms on EEG recordings. In: *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS).* (2017). p. 435–9. doi: 10.1109/CBMS.2017.83

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 02 August 2022
DOI 10.3389/fonc.2022.945362

Check for updates

OPEN ACCESS

EDITED BY
Thippa Reddy Gadekallu,
VIT University, India

REVIEWED BY
Zishao Zhong,
Guangdong Provincial Hospital of Chinese Medicine, China
Kai Yang,
Capital Medical University, China

*CORRESPONDENCE
Chengliang Yin
chengliangyin@163.com
Wei Zhao
34302603@qq.com
Haiwen Peng
penghaiwendoctor@sina.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

RECEIVED 16 May 2022
ACCEPTED 24 June 2022
PUBLISHED 02 August 2022

CITATION
Li W, Jin G, Wu H, Wu R, Xu C, Wang B, Liu Q, Hu Z, Wang H, Dong S, Tang Z-R, Peng H, Zhao W and Yin C (2022) Interpretable clinical visualization model for prediction of prognosis in osteosarcoma: a large cohort data study.
Front. Oncol. 12:945362.
doi: 10.3389/fonc.2022.945362

COPYRIGHT
© 2022 Li, Jin, Wu, Wu, Xu, Wang, Liu, Hu, Wang, Dong, Tang, Peng, Zhao and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Interpretable clinical visualization model for prediction of prognosis in osteosarcoma: a large cohort data study

Wenle Li[1†], Genyang Jin[2†], Huitao Wu[3], Rilige Wu[4], Chan Xu[5], Bing Wang[5], Qiang Liu[5], Zhaohui Hu[6], Haosheng Wang[7], Shengtao Dong[8], Zhi-Ri Tang[9], Haiwen Peng[10*], Wei Zhao[1*] and Chengliang Yin[11*]

[1]Department of Orthopaedic Surgery, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, Xianyang, China, [2]Department of Orthopedics, Hospital of People's Liberation Army of China (PLA), Wuxi, China, [3]Intelligent Healthcare Team, Baidu Inc., Beijing, China, [4]College of Information and Electrical Engineering, China Agricultural University, Beijing, China, [5]Clinical Medical Research Center, Xianyang Central Hospital, Xianyang, China, [6]Department of Spine Surgery, Liuzhou People's Hospital, Liuzhou, China, [7]Department of Orthopaedics, The Second Hospital of Jilin University, Changchun, China, [8]Department of Spine Surgery, Second Affiliated Hospital of Dalian Medical University, Dalian, China, [9]School of Physics and Technology, Wuhan University, Wuhan, China, [10]Orthopaedic Department, The Fourth Medical Center of People's Liberation Army of China (PLA) General Hospital, Beijing, China, [11]Faculty of Medicine, Macau University of Science and Technology, Macao, China

Background: Currently, the clinical prediction model for patients with osteosarcoma was almost developed from single-center data, lacking external validation. Due to their low reliability and low predictive power, there were few clinical applications. Our study aimed to set up a clinical prediction model with stronger predictive ability, credibility, and clinical application value for osteosarcoma.

Methods: Clinical information related to osteosarcoma patients from 2010 to 2016 was collected in the SEER database and four different Chinese medical centers. Factors were screened using three models (full subset regression, univariate Cox, and LASSO) *via* minimum AIC and maximum AUC values in the SEER database. The model was selected by the strongest predictive power and visualized by three statistical methods: nomogram, web calculator, and decision tree. The model was further externally validated and evaluated for its clinical utility in data from four medical centers.

Results: Eight predicting factors, namely, age, grade, laterality, stage M, surgery, bone metastases, lung metastases, and tumor size, were selected from the model based on the minimum AIC and maximum AUC value. The internal and external validation results showed that the model possessed good consistency. ROC curves revealed good predictive ability (AUC > 0.8 in both internal and external validation). The DCA results demonstrated that the model had an

excellent clinical predicted utility in 3 years and 5 years for North American and Chinese patients.

Conclusions: The clinical prediction model was built and visualized in this study, including a nomogram and a web calculator (https://dr-lee.shinyapps.io/osteosarcoma/), which indicated very good consistency, predictive power, and clinical application value.

KEYWORDS

osteosarcoma, SEER, multicenter study, nomogram, web calculator, prediction model

Background

Osteosarcoma, the most frequent primary malignancy of bone, accounting for approximately 35% of bone malignancy (1), originates from malignant mesenchymal cells (2), which produce osteoid and/or immature bone (3). Surgery combined with peri-operative chemotherapy is the current treatment while local therapy alone is insufficient (4). The presence or absence of metastases has become an important prognostic factor. Studies have shown that the 5-year survival rate for primary focus without metastases is more than 65% (5–7). Certain variables cannot explain the complicated survival rate. For diagnosis and treatment option, the American Joint Committee on Cancer (AJCC) system (8) and Enneking system (9) are popular. Factors of these systems can imply survival duration with treatment option roughly, but it is limited. A prediction model for survival is urgent for further prognosis prediction and instructive therapy selection (10, 11).

Osteosarcoma incidence remains low relative to other tumors (12). Therefore, a sufficient number of subjects are quite challenging. The Surveillance, Epidemiology, and End Results (SEER) database is an authoritative cancer statistics database in the United States that records morbidity, mortality, and incidence information for millions of patients with malignancies. Currently, although there have been relevant studies on osteosarcoma based on the SEER database, these prediction models showed a lower power (almost AUC < 0.8) or have no external data validation (13–15).

In this study, we built models based on osteosarcoma patients' data in the SEER database using three models, and the apt model was visualized. The validation data set from four different regional medical centers in China presented great power and credibility of the apt model. The nomogram and the web calculator were visualized, possessing good consistency and clinical application value.

Abbreviations: AJCC, The American Joint Committee on Cancer; AIC, Akaike's Information Criterion; AUC, area under the curve; DCA, decision curve analysis; MSE, mean squared error; ROC, receiver operating characteristic curve; SEER, The Surveillance, Epidemiology, and End Results.

Methods

Clinical information and selection criteria

SEER*STAT (version 8.3.5) software was used to extract data including patient demographic characteristics, clinicopathological treatment, and patient treatment (surgery, radiotherapy, and chemotherapy) information.

SEER data inclusion criteria were as follows: (1) primary malignant tumor of osteosarcoma with International Classification of Diseases of Oncology ICD-O codes 9180, 9181, 9182, 9183, 9184, 9185, 9186, 9187, 9192, 9193, 9194, and 9200; (2) SEER database after 2010 incorporated relevant metastatic site information and included patients diagnosed between 2010 and 2016; (3) osteosarcoma was the first and only primary malignancy; (4) complete clinical information, including age at diagnosis, sex, race, primary site, tumor size, tumor stage and grade, metastatic site, surgery, and whether radiotherapy and chemotherapy were administered; (5) diagnosis was from surviving patients and did not include cadavers; (6) complete follow-up information was available; and (7) known cause of death and survival time after diagnosis.

The multicenter data were obtained from four medical institutions in China: the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University, Liuzhou People's Hospital, and Xianyang Central Hospital. The follow-up period was more than 3 years. Three investigators were responsible for data acquisition at each institution during the survey period. Tumor size and stage were provided by the surgeon, and pathological grading was diagnosed by a senior pathologist at each hospital, or in case of uncertainty, confirmed by a pathologist at the Second Affiliated Hospital of Jilin University. Data were extracted by two of the three investigators, and data check was performed by the third

one. All data were checked for consistency and date was sorted using Microsoft Excel (Microsoft Excel, 2013, Redmond, USA).

Exclusion criteria were as follows: (1) incomplete clinicopathological and survival information; (2) unknown tumor size, stage, and race; and (3) vacant data.

Calibration of prediction model parameters and data baseline

Considering the characteristics of the SEER database and the multicenter study, we tried to unify the data standard. Three categories of race in SEER data were white, black, and other without specific subdivisions, while the race of real multicenter data from China was classified to "other". Treatment modality included surgery, chemotherapy, and radiotherapy, but the SEER database did not record treatment details; thus, it could only be classified as No (treatment) or Yes (treatment). Some patients were coded "999" on tumor size in the SEER database, which meant that their tumor size could not be assessed. To minimize data bias, we used x-tile to find the cutoff value of the data that can assess the tumor size, converting the tumor size from a continuous variable to a categorical variable.

Baseline tables were drawn for the modeling and validation group data, independent samples *t*-tests were used for continuous variables, and chi-square tests were used for categorical variables. Heat maps were plotted to show the frequencies and correlations between the parameters.

Selection of the prediction model

Three methods were used to screen variables in this study: (1) univariate Cox with $p < 0.05$ as a cutoff for screening variables and forest plot; (2) full subset regression to adjust for R^2 maxima to determine the best combination of variables; and (3) the LASSO regression and cross-validation to determine the combination of variables by the λ value while the mean squared error (MSE) was minimal.

The variables of the three methods were screened by using stepwise backward regression to achieve the minimum value of Akaike's Information Criterion (AIC). The models constructed by the three methods were compared by receiver operating characteristic (ROC) curves, and that with the largest area under the curve (AUC) was selected as the final model.

Survival analysis

In the prognostic analysis, Kaplan–Meier was used to estimate survival curves for each variable, and a log-rank test was used to determine the significant difference. Multivariate Cox regression analysis was used, and forest plots were drawn.

Development and visualization of prediction models

A nomogram was constructed using the parameters screened from the multivariate Cox results. For application convenience, a user-friendly web calculator was provided. Meanwhile, we built the decision tree.

Model validation and clinical application assessment

The actual and predicted probabilities were compared using calibration curves for the training and validation sets to evaluate the model consistency. The ROC of the validation set was plotted, and AUC was calculated to evaluate the prediction accuracy of the prediction model. Decision curve analysis (DCA) was used to evaluate the clinical application value.

Statistical analysis

Cutoff values were obtained by x-title software. Statistical methods and plotting, including *t*-test, chi-square test, LASSO, full subset regression, heat map, Kaplan–Meier, forest plot, nomogram, ROC curve, calibration plot, and DCA curve, were performed by R version 4.0.5. $p < 0.05$ was considered statistical significance.

Results

Continuous variables transformed into categorical variables

In Figure 1, the x-tile software calculated the optimal division of tumor size into the following groups: less than or equal to 95 mm, 95–127 mm, and more than 127 mm. Therefore, the continuous variable in tumor size was transformed into categorical variables in the three groups of ≤95, 95–127, and >127. Other patients coded "999" on tumor size were allocated to "Unable to evaluate".

Baseline data about SEER and multicenter data

Based on the SEER database, information was collected on all patients with osteosarcoma between 2010 and 2016, and according to inclusion and exclusion criteria, 1,144 patients were finally included, while a total of 112 patients were included in the Chinese multicenter data. Flowchart of data collection and analysis was shown in Figure S1.

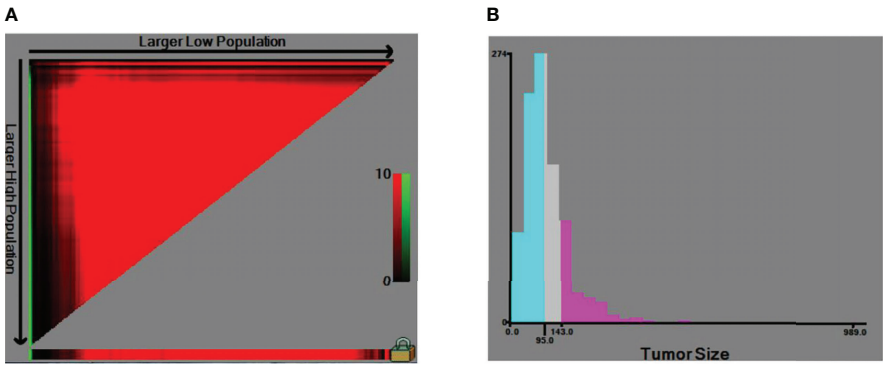


FIGURE 1
The cut-off of tumor size. **(A)** The x-tile software calculated the optimal division of tumor size. **(B)** Categorical variables.

Table 1 showed the demographic, clinicopathological, and treatment data characteristics of the SEER database versus the Chinese multicenter. Among the statistically significant differences between the two cohorts were race and chemotherapy. In the SEER data, Caucasians predominated, followed by blacks and other ethnicities. The multicenter data had a only Chinese population and a higher proportion of chemotherapy for osteosarcoma in China. No significant differences existed regarding other characteristics between the two groups (Table 1).

The heat map in Figure 2A showed the correlation between each parameter, and that in Figure 2B showed the frequency in each parameter. In Figure 2A, we could find moderate correlations for tumor size with T and stage group, race with category, and bone metastasis versus lung metastasis. In Figure 2B, the frequencies of each parameter were shown, and

TABLE 1 Baseline data table of the training group and the validation group.

Variable	Level	SEER data (training group N = 1,144)	Multicenter data (validation group N = 112)	p
Age, mean (SD)	NA	33.47 (24.26)	31.62 (24.88)	0.443
Survival times, mean (SD)	NA	29.91 (22.54)	30.10 (24.24)	0.933
Race (%)	Black	168 (14.7)	0 (0.0)	<0.001
	Other	116 (10.1)	112 (100.0)	
	White	860 (75.2)	0 (0.0)	
Primary site (%)	Axis bone	309 (27.0)	27 (24.1)	0.349
	Limb bone	738 (64.5)	79 (70.5)	
	Other	97 (8.5)	6 (5.4)	
Year of diagnosis (%)	2010	180 (15.7)	18 (16.1)	0.001
	2011	178 (15.6)	20 (17.9)	
	2012	197 (17.2)	16 (14.3)	
	2013	169 (14.8)	17 (15.2)	
	2014	193 (16.9)	14 (12.5)	
	2015	198 (17.3)	15 (13.4)	
	2016	29 (2.5)	12 (10.7)	
Laterality (%)	left	494 (43.2)	43 (38.4)	0.08
	Not a paired site	163 (14.2)	10 (8.9)	
	right	487 (42.6)	59 (52.7)	
Stage group (%)	I	182 (15.9)	17 (15.2)	0.816

(Continued)

TABLE 1 Continued

Variable	Level	SEER data (training group N = 1,144)	Multicenter data (validation group N = 112)	p
	II	521 (45.5)	49 (43.8)	
	III	44 (3.8)	7 (6.2)	
	IV	263 (23.0)	26 (23.2)	
	UNK stage	134 (11.7)	13 (11.6)	
T (%)	T1	388 (33.9)	38 (33.9)	0.294
	T2	523 (45.7)	46 (41.1)	
	T3	35 (3.1)	7 (6.2)	
	TX	198 (17.3)	21 (18.8)	
N (%)	N0	1,011 (88.4)	93 (83.0)	0.255
	N1	35 (3.1)	5 (4.5)	
	NX	98 (8.6)	14 (12.5)	
M (%)	M0	892 (78.0)	84 (75.0)	0.547
	M1	252 (22.0)	28 (25.0)	
Radiation (%)	No	999 (87.3)	104 (92.9)	0.119
	Yes	145 (12.7)	8 (7.1)	
Chemotherapy (%)	No	260 (22.7)	14 (12.5)	0.017
	Yes	884 (77.3)	98 (87.5)	
Tumor size (%)	>127	167 (14.6)	22 (19.6)	0.349
	≤95	552 (48.3)	47 (42.0)	
	59–127	235 (20.5)	21 (18.8)	
	Unable to evaluate	190 (16.6)	22 (19.6)	
Bone metastases (%)	No	1,044 (91.3)	102 (91.1)	0.981
	Unknown	47 (4.1)	5 (4.5)	
	Yes	53 (4.6)	5 (4.5)	
Surgery (%)	No	230 (20.1)	24 (21.4)	0.834
	Yes	914 (79.9)	88 (78.6)	

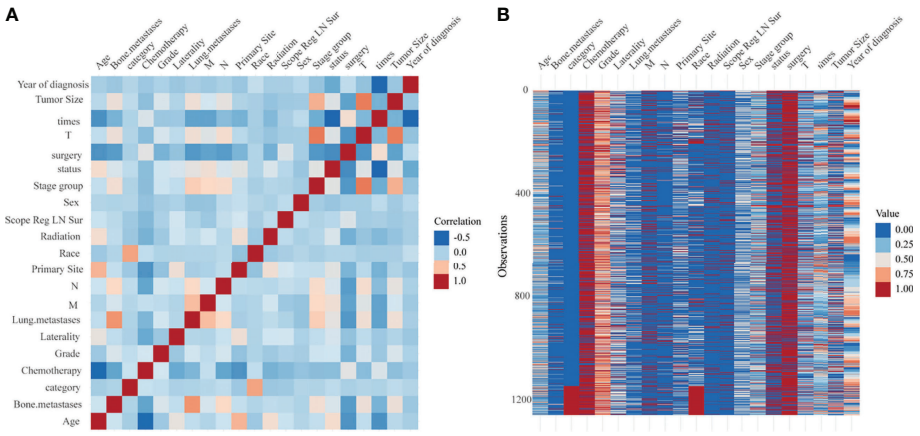


FIGURE 2
(A) Heat map of the correlation between each factor. **(B)** Heat map of the frequency in each factor.

the data distribution could be observed visually from the colors (Figure 2).

Univariate Cox regression

According to the results of the univariate Cox regression, the forest plot was drawn (Figure 3). According to the results of the univariate Cox regression forest plot, 14 variables ($p < 0.5$) were screened by univariate Cox regression, namely, age, primary site, grade, laterality, stage group, T stage, N stage, M stage, surgery, radiation, chemotherapy, bone metastases, lung metastases, and tumor size.

Full subset regression

The full subset regression was performed using the R packages' (leaps) regsubsets function to find the best combination according to the optimal subset regression model evaluation criteria, through adjusting the Marlowe's CP value to minimum, R^2 value to maximum, and Bayesian information criterion to minimum. The combination of variables was determined with the adjustment R^2 as criterion. Optimal full subset regression selected eight variables (age, grade, laterality, stage group, M stage, surgery, chemotherapy, and tumor size) (Figure 4).

LASSO regression and cross-validation

LASSO introduced the variable λ to find the most appropriate model. As λ increased, the regression coefficient β of each variable decreased, and some became zero, indicating that the variable contributed little to the model and could be eliminated. λ value determined which variables optimized the model, and the best λ value could be found using cross-validation (Figure 5A).

Figure 5B showed the Partial-likelihood deviance curve with Log(λ). The value (λ.min and λ.1se) was used to choose the good performance model in a minimum number of independent variables. Four combinations of variables (age, M, surgery, and lung metastases) were chosen *via* the LASSO regression.

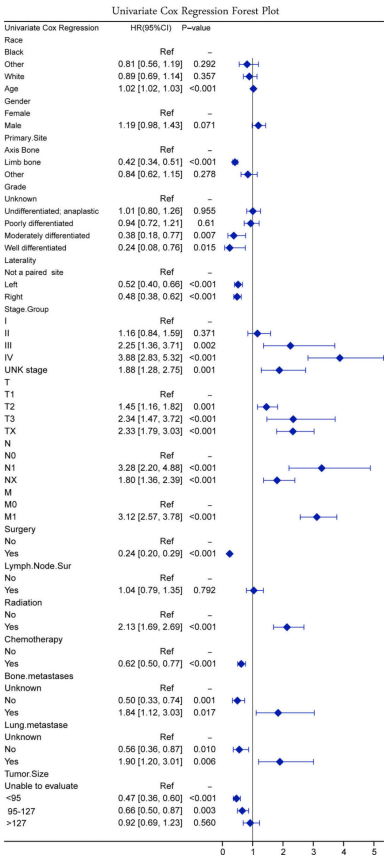


FIGURE 3
Forest plot on univariate Cox regression.

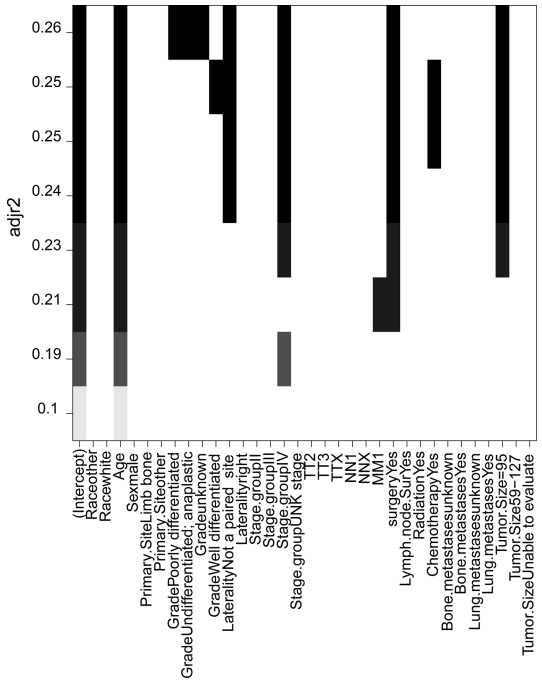


FIGURE 4
The combination of variables were determined with the adjustment R^2 as criterion in the full subset regression.

Multivariate Cox regression to determine the final model variables

A total of 14 variables were screened by univariate Cox regression. Based on adjusted R^2 maxima, eight variables (age, grade, laterality, stage group, stage M, surgery, chemotherapy, and tumor size) were screened by optimal subset regression (OSR). LASSO regression and cross-validation using a tuning factor (λ.1se) built an excellent model with a minimum number of four independent variables (age, M stage, surgery, and lung metastases).

The combinations of variables screened by each of the three methods were analyzed in a multivariate Cox model, and the final models of the three methods were determined using stepwise backward regression with minimum AIC values. After stepwise backward regression, eight variables were included in the univariate Cox (age, grade, laterality, M, surgery, bone metastases, lung metastases, and tumor size). Six variables were included in the optimal subset regression (age, grade, laterality, M stage, surgery, and tumor size). Four variables were included in LASSO regression (age, M stage, surgery, and lung metastases).

The AIC of the three models were 5,552.849 in univariate Cox, 5,570.204 in OSR and 5,611.193 in LASSO regression. ROC curves in three models were drawn at 1-year, 3-year, and 5-year survival. The models were evaluated by AUC values (Figure 6). The model constructed by univariate Cox was optimal with the largest AUC and the smallest AIC.

Survival analysis

The multivariate Cox forest plot showed that eight univariate Cox parameters were independent risk factors ($p < 0.05$, Figure 7).

Kaplan–Meier survival curves revealed that there was no significant difference in patient survival between the SEER data and the real Chinese multicenter data ($p > 0.05$, Figure 8A). Kaplan–Meier survival curves about all patients are presented in Figure 8B. Bone metastases were at higher risk than no bone metastases ($p < 0.05$, Figure 8C). Well-differentiated grade patients held longer survival ($p > 0.05$, Figure 8D). Kaplan–Meier survival curves almost overlapped in left and right laterality, showing no difference ($p > 0.05$, Figure 8E). Lung metastases were at higher risk than no bone metastases ($p < 0.05$, Figure 8F). M1 showed a lower survival rate than M0 ($p < 0.05$, Figure 8G). Patients with surgery showed higher survival rate than no surgery ($p < 0.05$, Figure 8H). The larger the tumor size of patients was, the shorter was their survival ($p < 0.05$, Figure 8I). The consistent results were proved in the validation cohort (Figure S2).

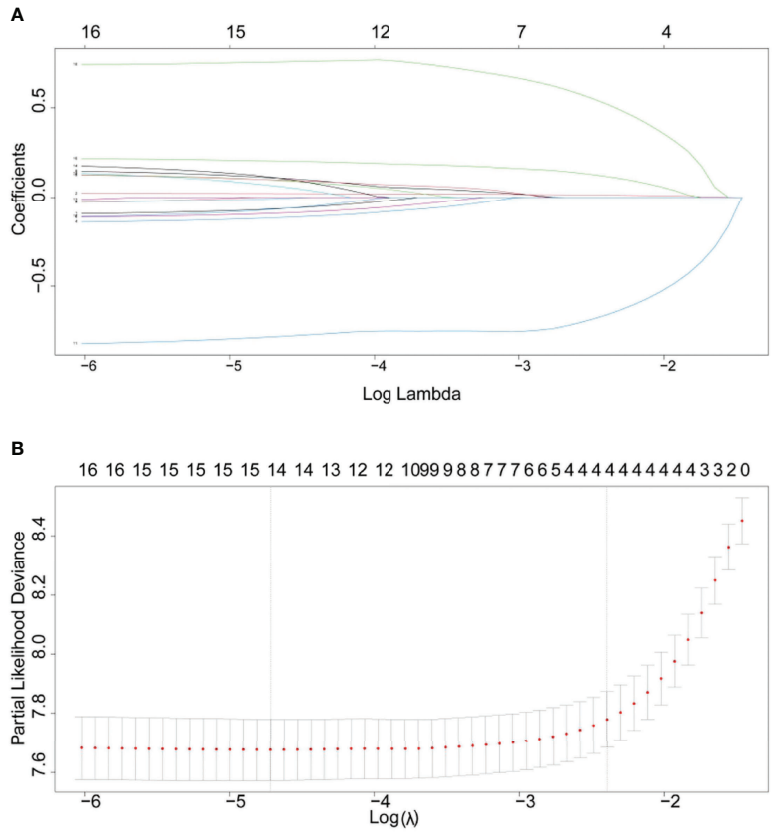


FIGURE 5
(A) LASSO coefficient profile. **(B)** Cross-validation for tuning parameter selection in the LASSO model.

Prediction model development

A nomogram is a method that allows quantification and visualization of Cox regression (16). The nomogram is evaluated by two methods: (1) Each variable is listed, and each sub-variable is quantified into a specific score. The cumulative scores of all variables are matched to the outcome scale to obtain predicted probabilities. (2) Web calculators or dynamic line graphs are developed to input specific variables and calculate the probability of an event. In this study, we constructed the nomogram using multivariate Cox variables (Figure 9A). Moreover, an online web calculator (https://dr-lee.shinyapps.io/osteosarcoma//) was designed to facilitate the user. A decision tree model was also provided as a supplement for the prediction model (Figure 9B).

Calibration chart and external receiver operating characteristic curve

The calibration chart was an assessment of how close the estimated risk of the line plot was to the actual risk. SEER data were applied for internal validation, and multicenter data were applied for external validation. The internal validation results (Figures 10A–C) and external validation results (Figures 10D–F) showed that the predicting outcomes were consistent with the actual outcome and the prediction model was well preformed in 1, 3, and 5 years. The ROC curves of the model were plotted in multicenter data. It proved the excellent predictive ability in 1, 3, and 5 years (AUC > 0.8, respectively) in Figure S3.

Risk score visualization and decision curve analysis

The risk score plots were used to visualize Cox survival risk models. Figure 11 could illustrate the risk factors heat map, the scatter plot of patients' status, survival time, and high/low risk group, in the training and validation groups, respectively.

As in Figure 12, in both the training and validation groups, there was no significant benefit for 1-year patients. In 3 years

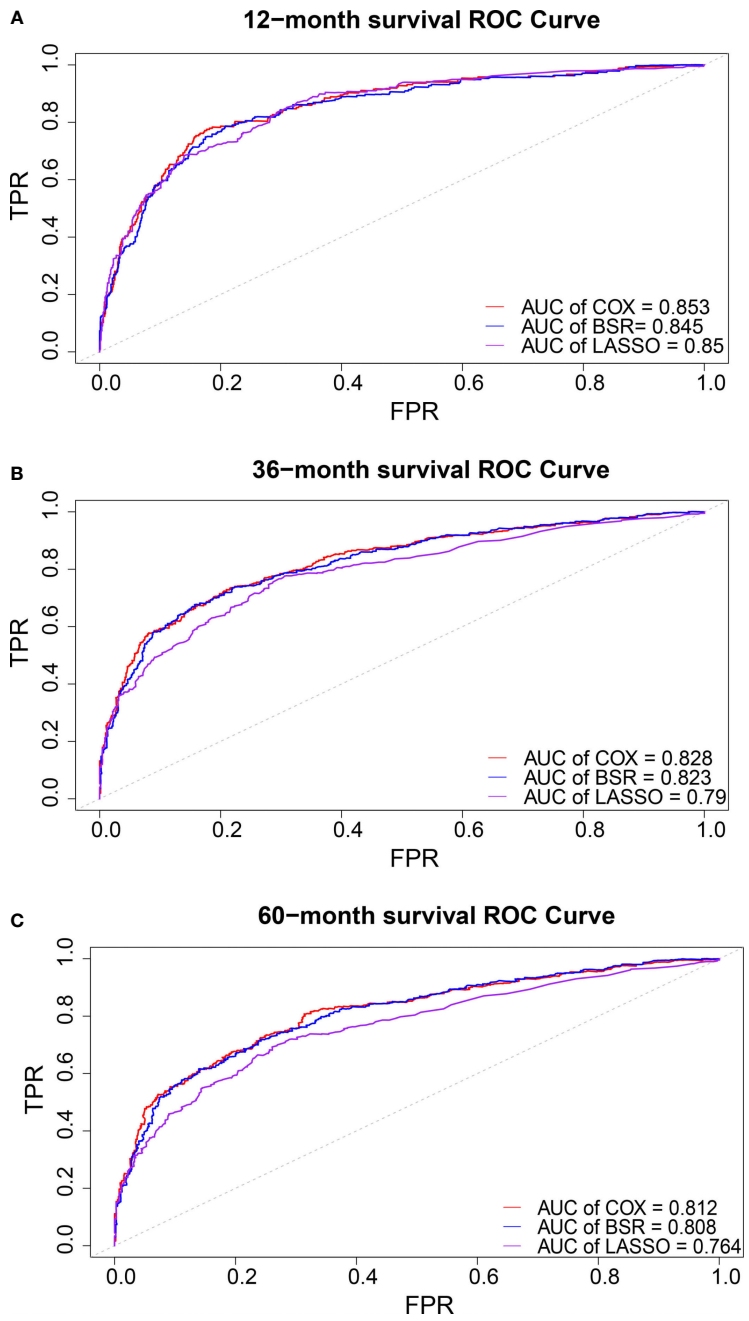


FIGURE 6
ROC cure in 1-**(A)**, 3-**(B)**, and 5- **(C)** years overall survival.

and 5 years, it was clear that the dashed line received a higher net benefit than the 1 year in both. Considering that osteosarcoma patients do not have a high mortality rate within 1 year of diagnosis, the prediction model developed in this study proved to have excellent clinical utility.

Discussion

Since the mid-1980s, with the standardization of treatment and the use of adjuvant chemotherapy, the 5-year survival rate for patients with osteosarcoma has arrived to approximately 65%

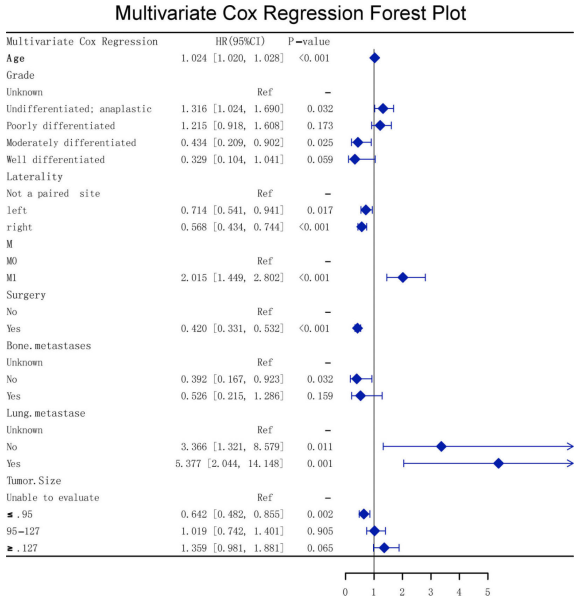


FIGURE 7
Multivariate Cox forest plot.

(17). No statistically significant differences were found between osteosarcoma patients from the United States and China, except in the ethnic distribution and the proportion of chemotherapy use (**Table 1**). All races were categorized as white, black, and other in SEER data, with other including Chinese. However, race was excluded as a predictor in three models (univariate Cox, full subset regression, and LASSO). In clinical practice, chemotherapy became a routine treatment option for metastases, especially concomitant lymphatic or vascular micrometastases (18). The proportion of chemotherapy in China was 87.5%, higher than that of the SEER database (**Table 1**). This might be related to the clinicians' preference and financial cause. Chemotherapy was a protective factor for patients with osteosarcoma in the univariate Cox results (**Figure 3**). However, the overall survival time from the multicenter data was not significantly different from the SEER database (**Figure 8A**). Proper indications for chemotherapy needed future research. This study had limitations since a retrospective study had bias in multicenter data collection, resulting in a higher proportion of chemotherapy.

In 2020, approximately 3,600 new cases with bone tumors and approximately 1,720 patients would die from the malignancy in the United States (19). Osteosarcoma with metastases clinically detected at initial presentation were approximately 20% of all osteosarcoma patients (17, 20). Approximately 30% of patients developed lung metastases within 1 year after diagnosis (21). Early detection of metastases could improve prognosis. Lung metastases had a strong correlation with bone metastases (M stage) and lymphatic metastases (N stage) (**Figure 2**). Bone metastases had a strong correlation with N. The mechanism of lymphatic metastasis from osteosarcoma has not been clear (22). Some studies found that osteosarcoma metastases disrupted the cortex, and the metastatic route might be through the lymphatic vessels of the synovial membrane and bursa (23). The incidence of lymphatic metastasis in patients with osteosarcoma did not exceed 5% in both SEER data and multicenter data, while lymphatic metastasis was a risk factor for patients with osteosarcoma in the univariate Cox results (**Figure 3**). Thus, we suggested that oncologists could not ignore the examination on lymph nodes. A nomogram has been constructed to predict distal metastases from osteosarcoma, which can be used as a method to screen for people at high risk of developing metastases (12). The first peak of mobility occurred at the age of 10–14 years, coinciding with pubertal growth (24, 25). Older patients may be less tolerant to treatment, and had a poorer prognosis. Osteosarcoma in an axial location showed poorer survival, and it was more difficult to completely resect focus due to location. Similarly, the larger tumor volume had poorer prognosis due to difficulty in complete resection, which was similar to previous studies (26).

AJCC (8) and Enneking (9) staging systems can only vaguely assess the clinical risk of osteosarcoma based on initial clinical features to help make treatment decision.

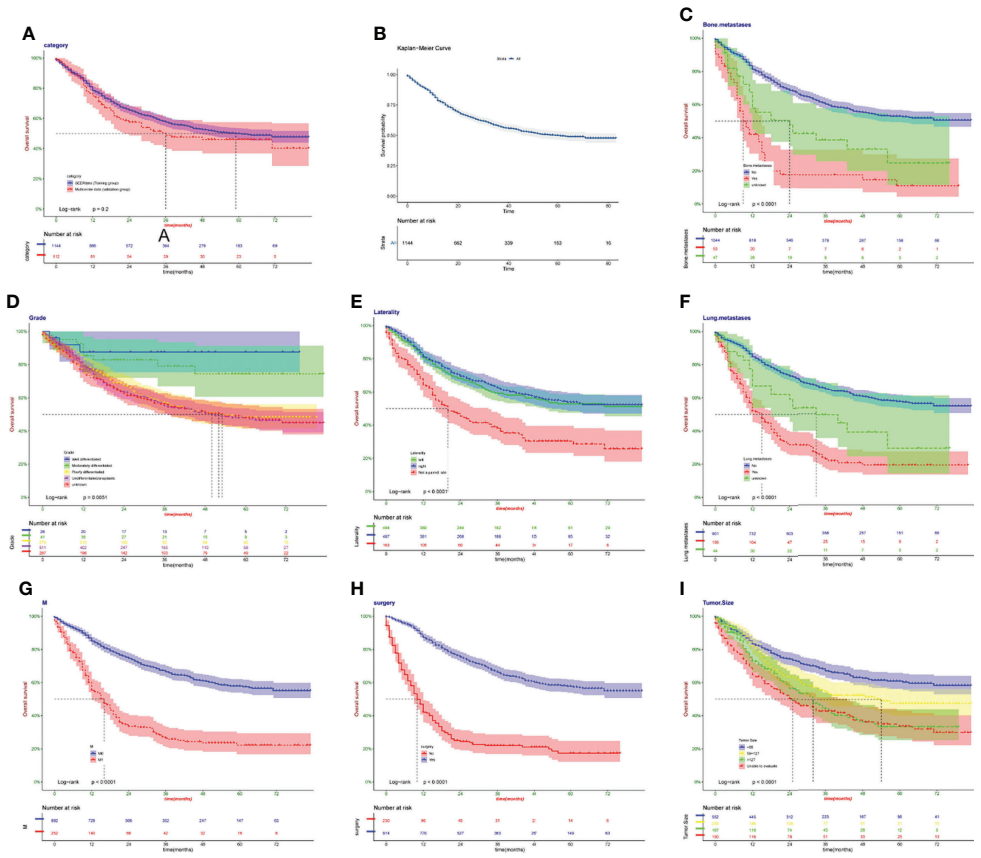


FIGURE 8
Kaplan−Meier survival curves in the training cohort. **(A)** The SEER data and the real Chinese multicenter data. **(B)** Patients in SEER data. **(C)** Bone metastases. **(D)** Grade. **(E)** Laterality. **(F)** Lung metastases. **(G)** M. **(H)** Surgery. **(I)** Tumor size.

Clinical prediction models are widely used today as tools for predicting the occurrence of specific events and estimating medical prognosis, especially in clinical oncology. Clinical prediction models generate probabilities of individual clinical events by integrating different predictor and decision variables, and their visualization and quantification advantages are also of great practical value in clinical practice (27). Most prediction models are developed based on logistic regression and Cox regression models. However, the full model equation remains difficult. In our study, three models (univariate Cox, full subset regression, and LASSO) were performed in the SEER database. Univariate Cox model and eight predicting factors (age, grade, laterality, stage M, surgery, bone metastases, lung metastases, and tumor size) were selected based on the minimum AIC and maximum AUC value. The model was further externally validated and evaluated for its clinical utility with data from four medical centers in China. ROC curves revealed good predictive ability (AUC > 0.8 in both internal and external validation, **Figure 6**).

A nomogram and web calculators were applied and visualized. Decision trees were provided as prediction model aids. A major advantage of the web calculator is that, compared to a rating scale or approximations calculated by the nomogram, the full model equation can be embedded in a backend web page, is more accurate in its calculation, and is more convenient to use. Web calculators can provide user-friendly graphical interfaces for complex mathematical models reducing the learning cost for users in today's world of smartphones and mobile networks. The nomogram is an effective quantitative method to assess risk and benefit and is widely used in clinical decision-making in a variety of diseases (28). In previous studies, several nomograms have been developed and validated to predict specific survival and overall survival in chondrosarcoma (29, 30).

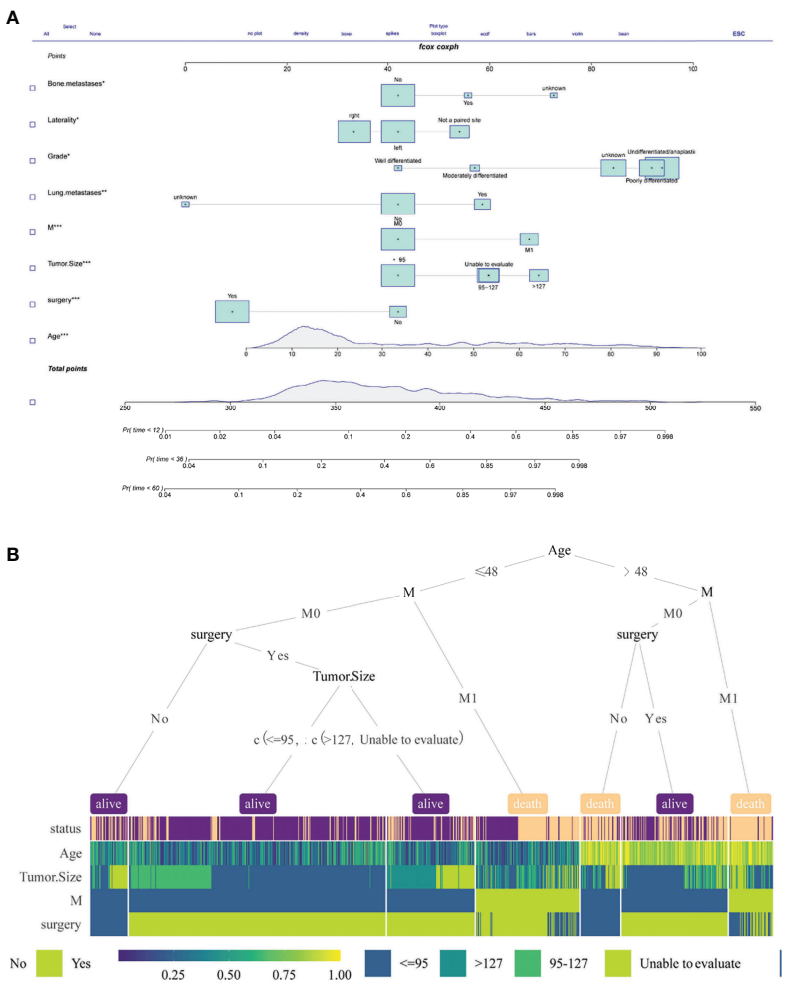


FIGURE 9
(A) A nomogram was constructed using multivariate Cox variables. **(B)** The decision tree of multivariate Cox variables.

In this study, a clinical prediction model to predict the overall survival of patients with osteosarcoma was developed to provide an objective reference for clinicians when making medical decisions. In clinical practice, lacking large-scale prognosis statistics for osteosarcoma patients in China, we chose the SEER database to develop the prediction model, and collected patient data from four medical centers in China to verify the feasibility of the model. In terms of clinical utility, the risk factor plots showed good stratification in both cohorts, effectively differentiating between high- and low-risk patient populations (**Figure 10**). The DCA displayed that both cohorts had better patient benefit from medical interventions in 3 years and 5 years. The 1-year model did not have a great net benefit, which may be related to the low mortality within 1 year (31).

Despite our efforts to refine the clinical prediction model, some limitations remained. (1) Training data (SEER database) to develop the prediction model were from North American patients, while the multicenter external data of China were tested for the model's predictive power. (2) The model was based on retrospective data and inevitably had inherent biases. (3) Previous studies showed that metastasis of osteosarcoma was associated with genes, metastatic mechanisms, proteins, and RNAs (32, 33). Since the SEER database did not contain relevant information, there was still room to improve the predictive power of the model.

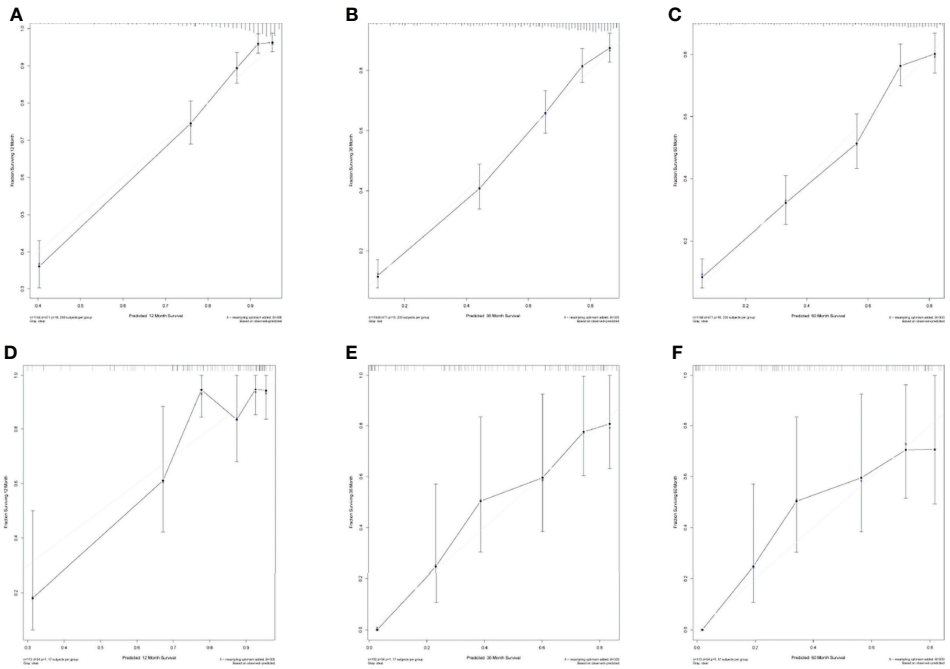


FIGURE 10
(A–C) Internal calibration diagram in 1 year, 3 years, and 5 years. **(D–F)** External calibration diagram in 1 year, 3 years, and 5 years.

Conclusions

In this study, based on the SEER database and data of osteosarcoma patients from 4 different regional medical centers in China, the model with the highest predictive ability was selected by three methods of screening model predictors, and the model was visualized for predicting the overall survival of osteosarcoma patients using three methods: nomogram, web calculator, and decision tree. The model was shown to have very good predictive power and consistency by both calibration plots and ROC curves. DCA demonstrated that the predictive model could provide greater benefit to patients. External validation results show that it still has predictive power and clinical use outside of North America.

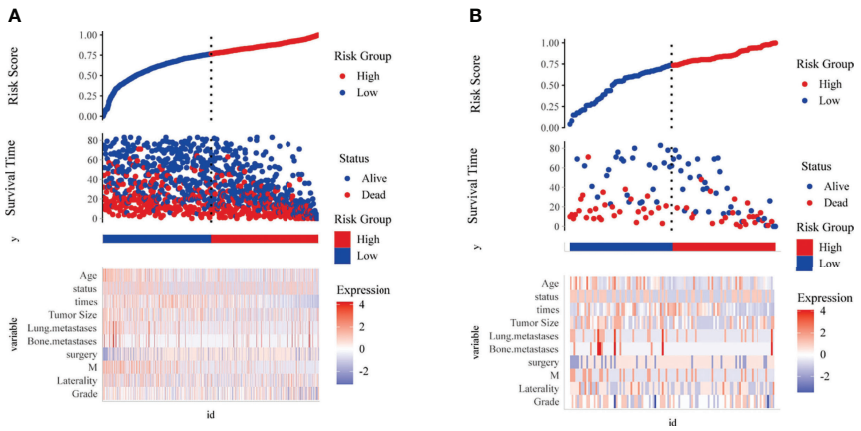


FIGURE 11
Risk score visualization. The scatter plot of risk score, the scatter plot of survival time and survival status for high and low risk, and the heat map of expression of key risk factors in the training group **(A)** and validation group **(B)**.

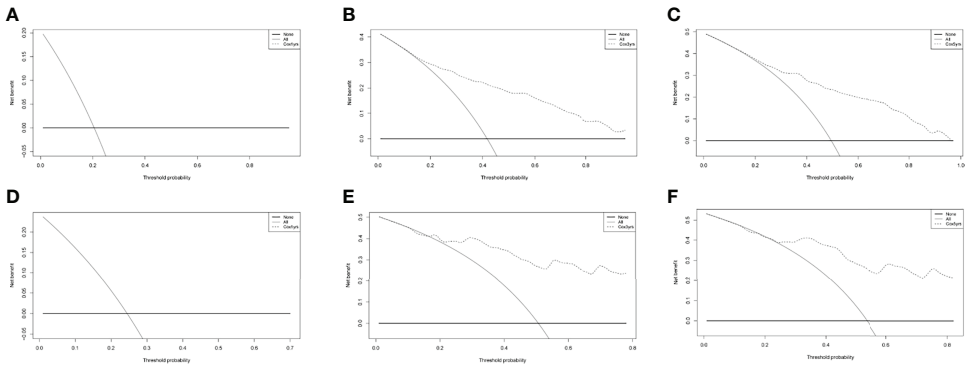


FIGURE 12
The DCA curves of the nomograms comparison for 1 year **(A)**, 3 year **(B)**, 5 year **(C)** in the training group, and for 1 year **(D)**, 3 year **(E)**, 5 year **(F)** in the validation group, respectively.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: SEER database within the article is public data set. The clinical information data of China's multicenter, analyzed during the current study are not publicly available for patient privacy purposes, but are available from the corresponding author upon reasonable request. Requests to access these datasets should be directed to CY, chengliangyin@163.com.

Ethics Statement

This study was exempted from Institutional Review Board approval, in view of the SEER's use of unidentifiable patient information. Due to the strict register-based nature of thestudy, informed consent was waived. The study of multicenter data was approved by the ethics review committee of the Second Affiliated Hospital of Jilin University, the Second Affiliated Hospital of Dalian Medical University,Liuzhou People's Hospital, and Xianyang Central Hospital(No. 20210021).

Author contributions

CLY, WZ, HWP: study conception and design; WLL, GYH: manuscript writing; HTW, RLGW, CX, BW, QL: literature review; all authors: data interpretation and discussion; all authors: final editing and approval of the manuscript in its present form.

Acknowledgments

We thank all individuals who took part in this research.

Conflict of interest

Author HW was employed by the company Baidu Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.945362/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Flow chart of data collection and analysis

SUPPLEMENTARY FIGURE 2
Kaplan–Meier survival curves in validation cohort. **(A)** the SEER data and the real Chinese multicenter data. **(B)** patients in multicenter data. **(C)** Bone metastases. **(D)** grade. **(E)** Laterality. **(F)** Lung metastases. **(G)** M. **(H)** surgery. **i** tumor size.

SUPPLEMENTARY FIGURE 3
External ROC curve in 1 year, 3 years, 5 years.

References

1. Jawad MU, Cheung MC, Clarke J, Koniaris LG, Scully SP. Osteosarcoma: Improvement in survival limited to high-grade patients only. *J Cancer Res Clin Oncol* (2011) 137:597–607. doi: 10.1007/s00432-010-0923-7

2. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: Data from the surveillance, epidemiology, and end results program. *Cancer* (2009) 115:1531–43. doi: 10.1002/cncr.24121

3. Marina N, Gebhardt M, Teot L, Gorlick R. Biology and therapeutic advances for pediatric osteosarcoma. *Oncol* (2004) 9:422–41. doi: 10.1634/theoncologist.9-4-422

4. Messerschmitt PJ, Garcia RM, Abdul-Karim FW, Greenfield EM, Getty PJ. Osteosarcoma. *J Am Acad Orthopaedic Surgeons* (2009) 17:515–27. doi: 10.5435/00124635-200908000-00005

5. Meyers PA, Schwartz CL, Krailo M, Kleinerman ES, Betcher D, Bernstein ML, et al. Osteosarcoma: A randomized, prospective trial of the addition of ifosfamide and/or muramyl tripeptide to cisplatin, doxorubicin, and high-dose methotrexate. *J Clin Oncol* (2005) 23:2004–11. doi: 10.1200/JCO.2005.06.031

6. Bacci G, Bertoni F, Longhi A, Ferrari S, Forni C, Biagini R, et al. Neoadjuvant chemotherapy for high-grade central osteosarcoma of the extremity. histologic response to preoperative chemotherapy correlates with histologic subtype of the tumor. *Cancer* (2003) 97:3068–75. doi: 10.1002/cncr.11456

7. Chou AJ, Geller DS, Gorlick R. Therapy for osteosarcoma: Where do we go from here? *Paediatric Drugs* (2008) 10:315–27. doi: 10.2165/00148581-200810050-00005

8. Greene F, Page D, Fleming I, Fritz A, Balch C, Haller D, et al. AJCC cancer staging manual. *JAMA J Am Med Assoc* (2010) 304:1726–7. doi: 10.1001/jama.2010.1525

9. William EF. Osteosarcoma. *Clin Orthopaedics Related Res* (1975) 111:2–4. doi: 10.1097/00003086-197509000-00001

10. Liu R-Z, Zhao Z-R, Ng CSH. Statistical modelling for thoracic surgery using a nomogram based on logistic regression. *J Thorac Dis* (2016) 8:E731. doi: 10.21037/jtd.2016.07.91

11. Li W, Dong S, Wang H, Wu R, Wu H, Tang ZR, et al. Risk analysis of pulmonary metastasis of chondrosarcoma by establishing and validating a new clinical prediction model: A clinical study based on SEER database. *BMC Musculoskeletal Disord* (2021) 22:529. doi: 10.1186/s12891-021-04414-2

12. Lu S, Wang Y, Liu G, Wang L, Wu P, Li Y, et al. Construction and validation of nomogram to predict distant metastasis in osteosarcoma: A retrospective study. *J Orthopaedic Surg Res* (2021) 16:231. doi: 10.1186/s13018-021-02376-8

13. He Y, Liu H, Wang S, Zhang J. A nomogram for predicting cancer-specific survival in patients with osteosarcoma as secondary malignancy. *Sci Rep* (2020) 10:12817. doi: 10.1038/s41598-020-69740-2

14. Fu P, Shi Y, Chen G, Fan Y, Gu Y, Gao Z. Prognostic factors in patients with osteosarcoma with the surveillance, epidemiology, and end results database. *Technol Cancer Res Treat* (2020) 19:1533033820947701. doi: 10.1177/1533033820947701

15. Qi L, Ren X, Liu Z, Li S, Zhang W, Chen R, et al. Predictors and survival of patients with osteosarcoma after limb salvage versus amputation: A population-based analysis with propensity score matching. *World J Surg* (2020) 44:2201–10. doi: 10.1007/s00268-020-05471-9

16. Balachandran VP, Gonen M, Smith JJ, Dematteo RP. Nomograms in oncology: More than meets the eye. *Lancet Oncol* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7

17. Ritter J, Bielack SS. Osteosarcoma. *Ann Oncol* (2010) 21 Suppl 7:vii320–5. doi: 10.1093/annonc/mdq276

18. Isakoff MS, Bielack SS, Meltzer P, Gorlick R. Osteosarcoma: Current treatment and a collaborative pathway to success. *J Clin Oncol* (2015) 33:3029–35. doi: 10.1200/JCO.2014.59.4895

19. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590

20. Kager L, Zoubek A, Pötschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary metastatic osteosarcoma: Presentation and outcome of patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol* (2003) 21:2011–8. doi: 10.1200/JCO.2003.08.132

21. Vormoor B, Knizia HK, Batey MA, Almeida GS, Wilson I, Dildey P, et al. Development of a preclinical orthotopic xenograft model of Ewing sarcoma and other human malignant bone disease using advanced *In vivo* imaging. *PLoS One* (2014) 9:e85128. doi: 10.1371/journal.pone.0085128

22. Dirik Y, Çınar A, Yumrukçal F, Eralp L. Popliteal lymph node metastasis of tibial osteoblastic osteosarcoma. *Int J Surg Case Rep* (2014) 5:840–4. doi: 10.1016/j.ijscr.2014.09.029

23. Edwards JR, Williams K, Kindblom LG, Meis-Kindblom JM, Hogendoorn PC, Hughes D, et al. Lymphatics and bone. *Hum Pathol* (2008) 39:49–55. doi: 10.1016/j.humpath.2007.04.022

24. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. *Cancer Treat Res* (2009) 152:3–13. doi: 10.1007/978-1-4419-0284-9_1

25. Simpson E, Brown HL. Understanding osteosarcomas. *JAAPA: Off J Am Acad Phys Assist* (2018) 31:15–9. doi: 10.1097/01.JAA.0000541477.24116.8d

26. Misaghi A, Goldin A, Awad M, Kulidjian AA. Osteosarcoma: A comprehensive review. *Sicot-j* (2018) 4:12. doi: 10.1051/sicotj/2017028

27. Sannino G, Orth MF, Grünewald TG. Next steps in Ewing sarcoma (Epi-) genomics. *Future Oncol (London England)* (2017) 13:1207–11. doi: 10.2217/fon-2017-0159

28. Li G, Tian ML, Bing YT, Wang HY, Yuan CH, Xiu DR. Nomograms predict survival outcomes for distant metastatic pancreatic neuroendocrine tumor: A population based STROBE compliant study. *Medicine* (2020) 99:e19593. doi: 10.1097/MD.0000000000019593

29. Gao Z, Lu T, Song H, Gao Z, Ren F, Ouyang P, et al. Prognostic factors and treatment options for patients with high-grade chondrosarcoma. *Med Sci monitor* (2019) 25:8952–67. doi: 10.12659/MSM.917959

30. Tang Z, Zhu R, Hu R, Chen Y, Wu EQ, Wang H, et al. A multilayer neural network merging image preprocessing and pattern recognition by integrating diffusion and drift memristors. *IEEE Trans Cogn Dev Syst* (2020). doi: 10.1109/TCDS.2020.3003377

31. Luetke A, Meyers PA, Lewis I, Juergens H. Osteosarcoma treatment - where do we stand? a state of the art review. *Cancer Treat Rev* (2014) 40:523–32. doi: 10.1016/j.ctrv.2013.11.006

32. Xiao Y, Zhao Q, Du B, Chen HY, Zhou DZ. MicroRNA-187 inhibits growth and metastasis of osteosarcoma by downregulating S100a4. *Cancer Invest* (2018) 36:1–9. doi: 10.1080/07357907.2017.1415348

33. Cao F, Kang XH, Cui YH, Wang Y, Zhao KL, Wang YN, et al. [Upregulation of PLOD2 promotes invasion and metastasis of osteosarcoma cells]. *Zhonghua zhong liu za zhi [Chinese J Oncol* (2019) 41:435–40. doi: 10.3760/cma.j.issn.0253-3766.2019.06.007

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 09 August 2022
DOI 10.3389/fpubh.2022.924025

Check for updates

OPEN ACCESS

EDITED BY
Thippa Reddy Gadekallu,
VIT University, India

REVIEWED BY
Vinodhini Mani,
SRM Institute of Science and Technology, India
Sri Vidhya,
Sathyabama Institute of Science and Technology, India

*CORRESPONDENCE
Chenliang Deng
dengclkd@163.com

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 20 April 2022
ACCEPTED 02 May 2022
PUBLISHED 09 August 2022

CITATION
Deng C, Yu Q, Luo G, Zhao Z and Li Y (2022) Big data-driven intelligent governance of college students' physical health: System and strategy. *Front. Public Health* 10:924025. doi: 10.3389/fpubh.2022.924025

COPYRIGHT
© 2022 Deng, Yu, Luo, Zhao and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Big data-driven intelligent governance of college students' physical health: System and strategy

Chenliang Deng[1]*, Qiaoyan Yu[2], Ganglin Luo[3], Zhangzhi Zhao[1] and Yuchao Li[1]

[1]Sports Department, University of Electronic Science and Technology of China, Chengdu, China, [2]School of Gymnastics, Chengdu Sport University, Chengdu, China, [3]Department of Military Education and Training, Police Academy of the Armed Police, Chengdu, China

With the development of information technology, the application of a new generation of information technologies, such as big data, Internet Plus, and artificial intelligence, in the sports field is an emerging, novel trend. This paper examined the relevant research results and literature on physical education, computer science, pedagogy, management, and other disciplines, then used a self-made questionnaire to investigate the physical health status of Chinese college students. The big data were subsequently analyzed, which provided a scientific basis for the construction of an intelligent governance system for college students' physical health. Intelligent devices may be used to obtain big data resources, master the physical sports development and psychological status of college students, and push personalized sports prescriptions to solve the problems existing in college students' physical health. Research shows that there are four reasons for the continuous decline in Chinese college students' physical health levels. These are students' lack of positive exercise consciousness and healthy sports values (85.43%), a weak family sports concept and lack of physical exercise habits (62.76%), poor implementation of school sports policies (55.35%), and people's distorted sports value orientation (42.27%). Through the connecting effect of data, we can bring together the positive role of the government, school, society, family, and students so as to create an interlinked impact to promote students' physical health. The problems of insufficient platform utilization, lack of teaching resources, lagging research, and insufficient combination with big data in the intelligent governance of physical health of Chinese college students can be solved by building an intelligent governance system of physical health. Such a system would be composed of school infrastructure, data resources and technology processing, and intelligent service applications. Among these, school infrastructure refers to the material foundation and technical support. The material foundation includes perceptions, storage, computing, networks, and other equipment, and the technical support includes cloud computing, mobile Internet, the Internet of Things, artificial intelligence, and deep learning. Data resources refer to smart data, such as stadium data, physical health management data, and students' sports behavior data, which are mined

from data resources such as students' physical development, physical health, and sports through big data technology and intelligent wearable devices. Intelligent managers provide efficient, intelligent, accurate, and personalized intelligent sports services for college students through data resource value mining, venue space-time optimization, health knowledge discovery, sports prescription pushes, etc. Finally, we put forward the development strategy for further deepening and improving the big data-driven intelligent governance system for college students' physical health. The intelligent governance system of physical health driven by big data and its development strategy can not only accurately guide and improve the physical health level of college students but also realize integrated teaching inside and outside physical education classes.

KEYWORDS

information technology, big data, information physical education, physical health of college students, intelligent governance of physical health

Introduction

Physical health is the basic right and the first wealth of college students and, more importantly, the basic premise of their healthy study, work, and life in the future. Therefore, how to improve the physical health level of college students is a very meaningful and valuable work pursued by the country, school, society, family, and students.

During the past 70 years, since the founding of the People's Republic of China, the party and the state have remained concerned and attached great importance to the health of students. To encourage and promote students to actively participate in physical exercise and enhance students' physical fitness, the relevant competent departments developed a series of systems in different periods, such as the "Labor Hygiene System" (1954), "National Physical Exercise Standards" (1975), "Student Physical Education Qualification Standard" (1990–1991), and "Sports Examination Methods for Junior High School Graduates" (1997). In 2002, the "Standards for Students' Physical Health" was piloted in schools across the country and has been implemented for more than 10 years. In 2014, the Ministry of Education issued a new "Standards for Students' Physical Health," which required that "all schools carry out physical health tests covering students of all grades every school year and stipulates that students whose physical health test score is <50 points will be treated as lacking completion or having incomplete study and unqualified schools will be rejected in the evaluation." In 2016, the Opinions on Strengthening School Physical Education and Promoting the All-round Development of Students' Physical and Mental Health contended that "we should carry out strategic, forward-looking, and applied project research aimed at improving students' physical health, teaching quality, after-school training, and sports cultural level through multiple channels so as to improve the overall level of school physical education work." In 2018, President Xi Jinping stressed at the National Education Conference that "health comes first in education, and physical education classes should be provided to help students enjoy fun, enhance their physical fitness, improve their personality, and temper their willpower." In 2020, the Opinions on Comprehensively Strengthening and Improving School Physical Education in the New Era clearly pointed out that "under the guidance of Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, by fully implementing the party's education policy, adhering to the socialist direction of school running, and taking moral education as the fundamental and socialist core values as the guidance, we aim to serve the all-round development of students and enhance their comprehensive quality, adhere to the education concept of health first, and promote the coordinated development of cultural learning and physical exercise for young people." In the same year, the "Notice on Deepening the Integration of Sports and Education to Promote the Healthy Development of Teenagers," which was jointly issued by the General Administration of Sport of China and the Ministry of Education, also suggested to "establish the education concept of health first, then open full physical education for all students, to help students enjoy fun in physical exercise, enhance their physical fitness, improve their personality, temper their will, realize the civilization of their spirit, and savage their body." In 2021, the "Opinions on Comprehensively Strengthening and Improving School Health and Health Education in the New Era" suggested to "implement the fundamental task of moral education, adhere to the education concept of health first, and comprehensively improve students' health literacy into a high-quality education system as an important goal and evaluation standard of school education. We will deepen the reform of health education in schools; promote students' physical and mental health, develop a healthy lifestyle; and

train them to be socialist builders and successors who are well-developed morally, intellectually, physically, aesthetically, and laboriously." The implementation of a series of national policy documents and systems has pointed out the direction for the development of physical education in schools, and it is of great significance to further strengthen the physical education of young students, enhance their physique, and vigorously promote quality-oriented education.

At present, domestic experts and scholars' research on big data sports is still in the initial stage of exploration, and some research results have been accumulated. The research directly related to this paper mainly includes four aspects, one of which is research on big data in sports. Since 2013, big data in sports has become a hot topic widely discussed. In 2015, the five trillion sports industry's goal was to attract Internet giants to break into big data of physical layout, but the research on theory and practice is limited and only a small number of researchers' findings have focused on big data in sports, new ecosystems, forward-looking analysis, sports communication, etc. The second aspect is the research on the sports industry against the background of big data. Representative research in this field includes work by Tong et al. (1), Dong et al. (2), Yu (3), Xie et al. (4), and other documents and materials. This research has mainly focused on the development of the sports industry in the era of big data, including opportunities and challenges, paths, and financial risk prevention and control as well as facility operation, operation for sporting events, sports marketing, sports industry integration, and so on. The third aspect is the research on physical education in the era of big data. Currently, there are only a few studies, mainly those of Zhang and Sun (5). The research results mainly discuss the application of big data in school physical education, the reform of physical education teaching methods, the change of physical education learning methods, and the precision teaching mode of physical education from the aspects of appeals, trends, and action. Fourth is the study of intelligent sports. Representative research results in this aspect are concentrated in the areas of intelligent sports classrooms (6), intelligent sports stadiums (7), intelligent sports services (8), and intelligent governance of public sports services (9). Experts and scholars have conducted their research from the perspectives of the Internet, big data, the mobile Internet, cloud platforms, and artificial intelligence. Overall, the related research results for this research laid the foundation and provide a theoretical reference. However, available big data technology research to improve the physical health of university students is limited; the qualitative-quantitative combination of sports science and medicine is lacking; the combination of computer science, management, sociology, and related influencing factors of college students' physical health in comprehensive research is unexplored; and there is no research on college students' physical health service system driven by big data.

With the in-depth development of information technology, big data has brought new development opportunities to all walks of life, and school physical education is also currently facing many opportunities and challenges. The integrated development of school physical education and information technology has become a core task and urgent demand to solve the existing problems of school physical education in China. At present, the development of school sports informatization in China is poor, lacking both talents and technology, which does not match the current national strategic environment. Given the current research and development situation, how to integrate college students' physical health and college sports resources with computer technology, physical health services, and other elements remains challenging. Equally daunting is how to build a big data-driven college students' physical health service system for management coordination, decision support, intelligent governance, and the construction and development of college students' intelligent sports venues. Therefore, based on the challenges of college students' physical health under the current big data environment and the relevant research status of intelligent governance, this study starts from the key integration of advanced information technology, such as big data, and the development of student's physical health. It carries out innovative research in theory, technology, application, evaluation, and other aspects, and improves the level of college students' physical health through data accuracy technology to build an intelligent governance system for college students' physical health based on big data.

Methods

Based on the theoretical and practical knowledge of school physical education, computer science, pedagogy, management, and other disciplines, this paper studied the research results and available literature on college students' physical health at home and abroad and used existing theoretical and practical results to support and build the theoretical framework and methodology of this paper.

By using self-compiled questionnaires to investigate the physical condition of college students and through big data analysis and sorting, this study provides a scientific basis for building an accurate system for the improvement of college students' physical health. To collect and assess students' physique, health, and influencing factors, the head of the department in charge of local sports and sports science, management, and computer science or a scholar of related fields visited the relevant domestic school to conduct a full investigation.

The influencing factors of this study are determined by a factor analysis method. Based on the contents obtained from the literature discussion and interviews with experts and scholars, questionnaires were designed according to the preliminarily constructed influencing factors of college students' physical health; three rounds of expert surveys were undertaken to

establish the influencing factors. In the study of students' physique health service systems, determination by wisdom is one of the most widely used methods in social science research employing the Delphi method. Through literature review, interviews of experts, and based on the content according to the preliminary build of the students' physique health service system, designed the questionnaire, performed three rounds of expert investigation, and established the system.

We grasped students' physical sports development, and psychological conditions by using emerging information technology that pushed personalized teaching and sports prescriptions, and addressed the existing problems of college students' physical health by using big data technology. We summarized and studied while providing timely feedback and revision of the plan, and performed the experiment laid out in this study.

Results

Analysis of reasons for the decline in Chinese college students' physical health

In recent years, the physical health level of Chinese college students has continued to decrease due to the effect of subjective and objective factors of the government, schools, society, families, and students.

Subjective factors of declining physical health of college students

According to the survey results, at present, 85.43% of students in Chinese universities lack active sports consciousness and correct values. In other words, they do not know why they want to exercise. As Mao Zedong said in the vernacular translation of the Study of Sport: to deal with a matter is reflected in behavior; we must first start to be interested in or like it. In particular, we must first have a specific understanding of the matter and why such wisdom is necessary and must clearly understand the specifics and why, which is the embodiment of self-consciousness. Many college students do not know the benefits of sports for themselves, either very well or not at all. In short, the stimulation of sports potential is inadequate, so there is no emotion to mobilize sports interest. Those students who can study tirelessly and assiduously think that the relationship between sports and learning is not close, and their survival is not based on sports. Some even think that sports will delay their cultural learning and other activities. Therefore, sports self-consciousness is resulting in worrying physical health. This portion of the reason for non-participation should be blamed on students' inability to deeply understand the benefits of sports. However, teachers do not know how to enlighten students or how to work hard ideologically, which also accounts for half of the responsibility. On the other hand, students lack correct sports values. They are embarrassed to exercise and consider people who can exercise stupid. This is the big reason why people do not like sports. Gentle and slow people look very good and are also respected by society. What does a person who suddenly opens their arms and feet, stretches their limbs, and bends their body look like? Shouldn't this be considered particularly strange? So, there are people who know that the body must move and want to take action but don't take action. There are also people who can exercise together but can't exercise alone, and then there are those who can exercise in their own home but can't exercise in places with many people (10). In short, the subjective reason for the decline of college students' physical health is that they are self-conscious and shy, do not form the habit of physical exercise, and do not establish a correct concept of physical education.

Objective factors of declining physical health of college students

One objective factor at play is the concept of family sports and exercise habits, which accounts for 62.76%. Family is the main place where students grow up and live and also the important base of students' enlightenment and education (11). In a sense, the concept of family sports, sports atmosphere, and sports habits will directly affect the enthusiasm of students in participating in physical exercise. Specifically, if the family has good sports traditions and a physical exercise atmosphere and parents generally like to participate in sports fitness activities, students will naturally fall in love with sports fitness activities in college. These individuals tend to experience the passion and fun of sports in a family atmosphere. On the other hand, if the family does not have the atmosphere and tradition of sports activities, their participation in sports activities will inevitably be less. In effect, the concept of "preconception gives priority to" decides the outcome. Compared to school education, family education is more likely to stimulate students' enthusiasm and participation. Family intimacy, tacit understanding, and other factors determine the important position of family education in cultivating students' good physical exercise habits (12). According to a 2019 survey report, 64.8% of Chinese parents said physical health is always the top priority, while 75.9% of parents of kindergarteners and >50% of parents in other learning stages support physical education. In addition, 73.6% of parents believe a lack of exercise is the main reason for college students' poor physical fitness. However, although the importance of physical development is recognized, in the actual home environment, the lack of momentum is very serious. A family's poor sports atmosphere and lax education are common phenomena. On the one hand, parents exercise an average of 2.5 h a week, and >60% exercise with their children no more than once a week. On the other hand, 66.8% of parents have adopted a relaxed educational approach to cultivate their children's sports habits,

focusing on their children's interests rather than discipline. Only 1/3 of parents adopt a strict physical exercise regimen. The decrease in exercise time and frequency is directly reflected in the physical condition of students, resulting in the continuous decline of students' physical health levels. Parents rate their children's physical health as passing, not good. Parents of high school students awarded the lowest scores for their children's physical health.

Second, the school sports policy implementation deviation accounts for 55.35%. According to the survey results, most colleges and universities have problems with the implementation of school sports policies, be it in its scale, insufficient depth, insufficient strength, or false implementation methods. In fact, most schools have not really implemented relevant sports policies, which has led to a sharp decline in the physique of Chinese teenagers to a certain extent. Additionally, the school does not pay enough attention to the cultivation of core literacy regarding physical fitness. The study of cultural courses occupies a large portion of the college students' spare time and increases with the advancement in grade, resulting in the lack of exercise time among college students. Nearly 67.6% of college students exercise no more than twice a week. The above problems are caused by the issues in school sports policies, their weak implementation, and the lack of an effective supervisory mechanism. Therefore, we should improve and optimize the level of policy formulation, enhance capacity for policy implementation, identify and support policy implementation objectives, provide resources and the environment for policy implementation, and have a monitoring and oversight system for policy implementation (13).

Finally, the influence of distorted sports value orientation of social groups accounts for 42.27%. In China, although people know they should be active in sports and believe that promoting sports can make a country strong, they tend to pay more attention to cultural education than physical education. Such ideas and actions lead most people to hover between welcoming and rejecting sports. So, most people do not really like sports. Second, most of today's educators are not good at sports. They don't know much about sports and may have only heard of the word "sports," so when they come to teach sports, they start from a dishonest point of view, and they can't or won't do anything about it, so they weaken students' curiosity about sports. In addition, most people who teach physical education have shallow knowledge. They only know a few sporting events, and they may not be proficient in them. They lack in-depth knowledge and only their mechanical movements are seen by others every day. Therefore, many think that sportspersons are uneducated. Indeed, only focusing on the "body" of sports without also focusing on "education" or only paying attention to the form but not the substance will not do for a long time, and the fact of today's school sports is the same.

Current situation of intelligent management of physical health of Chinese college students

At present, information technology closely links the world promoting the process of sports informatization and development of sports worldwide (14). The intelligent management of physical health has become an important way to improve the physical health level of college students and an important means to promote the intelligent services of college students' physical health. However, there are still many problems in the intelligent governance of college students' physical health, such as insufficient platform utilization, lack of resources, lagging research, and a low governance level.

Insufficient use of intelligent governance platforms for college students' physical health

According to our survey, although some Chinese universities have established an information platform for student physical health monitoring systems, the platform is only a decoration and has not been really used. Both the Ministry of Education and the General Administration of Sports have information technology centers, which are mainly responsible for monitoring and have been specifically set up to relay the macro-management and scientific decision-making efforts of ministries and commissions. Despite the establishment of sunshine sports and national fitness platforms, due to its outdated model lagging in products and services, their benefits are poor. How to reasonably combine school sports with informatization, build an intelligent governance platform for physical health in colleges and universities, and make full use of this platform to guide students' scientific sports and healthy life is an urgent problem that needs to be solved.

Lack of intelligent physical health service teaching resources for college students

For a long time, physical education teaching in colleges and universities has had a single structure, outdated content, traditional methods, and insufficient interactions. Teachers are the absolute owners of knowledge, and students can only learn from teachers in class. Therefore, the lack of communication and sharing in teaching makes it difficult to carry out educational and teaching activities creatively. The construction of physical health intelligent governance systems driven by big data has had great influence and significance. On the one hand, information-based physical education teaching resources can break through the traditional narrow teaching methods, letting the life experiences of teachers and students enter the physical education teaching process and rendering the classroom truly active. On the other hand, it can also change the status of learners in the physical education curriculum from passive cognitive audience

members to knowledge-builders so as to transfer students' learning interests and play a positive role. At the same time, it can also broaden teachers' educational horizons, alter teachers' teaching concepts, and obtain the possibility of interaction and resource-sharing with various sports teaching resources, especially teaching material resources. The possibility of mutual cross-transformation of sports teaching resources inside and outside the school will also be increased, and an integrated sports teaching method inside and outside the curriculum will really be realized. Although many enterprises, colleges and universities, and individual teachers have carried out corresponding sports teaching resource construction projects by using big data, and networks, most have employed single sports material resource databases. At the same time, due to the various selection and use environments of sports materials, they sometimes have no way to start within their own schools or individual teachers in the selection process, and even need to be modified again; therefore, teachers must spend significant amounts of time and spiritual resources to do this kind of thing well. Sometimes, it goes against the original design and even impacts the improvement of school physical education teaching quality, which has no obvious effect on the improvement of college students' physical health. Colleges and universities should make full use of information technology, integrate physical education teaching resources, integrate network physical education courses, build high-quality and rich physical education teaching resources, and attain diversified physical education teaching processes and space. They must enable teachers to obtain physical education teaching resources they want in a limited time frame and encourage students to adopt exercise habits through an effective autonomous learning mode so as to finally achieve the purpose of improving their health (15).

The research on intelligent physical health services for college students lags behind

At present, there is still a lack of sports informatics specialty knowledge in China. The research on physical health and intelligent governance is mostly interdisciplinary research, while at the application level, it focuses on sports products on the market, mainly forming a new ecological environment of the Internet industry in the form of a platform, including mobile terminal applications such as sports fitness and running programs. Because China's sports intelligent governance started late and lacks relevant disciplines, there are those who understand sports but do not understand informatization, and those who understand informatization do not understand sports. Moreover, these mobile terminal products and information technology research fields are not well-integrated, and the technical content of products is generally not high. They mainly make profits through content and capital movement. Based on the Internet of Things, artificial intelligence competition training, physical education, physical health monitoring, and other research fields, there is still a large space to fill at present.

The level of intelligent governance of college students' physical health is low

According to the survey results, college students' physical health management services still lack the support of big data, mainly in three aspects. First, the college sports intelligent governance system is not enough. Only the venues in China's college sports intelligent governance at present have the "Internet of Things," while construction at the infrastructure level remains focused on local support of facilities for activity service processes. To realize big data-driven college sports intelligent governance in China, we must establish a systematic and overall intelligent governance system based on intelligent technology and infrastructure. In addition, Chinese universities pay more attention to static information, while real-time and dynamic stream data information has not been widely used, and this data information is often necessary to support the school's intelligent movement behaviors. Second, the current results of big data analysis of college students' physical health have not been systematically organized and managed. In its current form, the existence of large quantities of college students' physical health information data increases the difficulty of obtaining big data analysis information from relevant institutions, which seriously interferes with the reasonable screening, identification, and mining of information data. Third, the current college students' physical health service system structure has been unable to meet the requirements of big data-driven intelligent development. At present, the collection, management, storage, and big data analysis of college students' physical health information and the school decision-making system have been unable to support decision-making through big data technology and cannot realize intelligent school physical education and intelligent physical health services.

Construction of big data-driven intelligent governance system for college students' physical health

From the perspective of being big data-driven, this paper explores the development of a college students' physical health intelligent governance system. According to information ecological chain theory and guided by college students' physical health, a big data-driven college students' physical health intelligent governance system has been formed that can integrate infrastructure, health data resources, and business applications with students; explore the value of data intelligence in college students' physical health data resources by using big data analysis technology; provide learners with efficient, intelligent, accurate, and humanized intelligent sports consulting services;

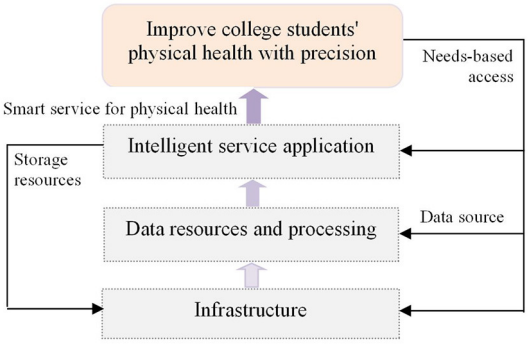


FIGURE 1
Data-driven intelligent governance system architecture of college students' physical health.

and realize integrated physical education inside and outside the classroom. Intelligent wearable devices are used to achieve physical fitness testing, sports monitoring, disease and health management, and dynamic adjustment of physical education teaching content. In combination with the physical health management system, a home-school linkage mechanism is established to master students' physical sports development, and psychological conditions with the use of emerging information technology. Personalized teaching and exercise regimens are pushed to truly achieve individualized teaching and accurate guidance and improvement of college students' physical health.

The data-driven intelligent governance system for college students' physical health includes three levels: school infrastructure, data resources and processing technology, and intelligent service application (see Figure 1). Among these, school infrastructure refers to the material foundation and technical support. The material foundation includes sensing, storage, computing, networks, and other equipment, and the technical support includes cloud computing, mobile Internet, the Internet of Things, artificial intelligence, and deep learning. Data resources are those available through big data technology and intelligent wearable devices about students' physical development, physical health, and movements at sports venues data mining. Data resources include physical health management data and student behavior data such as wisdom and are made available by machine learning, artificial neural networks, data visualization, and social network analysis means and methods for data processing to provide managers with intelligence data services. Smart administrators provide efficient, intelligent, accurate, and personalized smart sports services to college students through data-driven procurement (exercise information, good exercise methods), venue space-time optimization (field exercise experience), health knowledge discovery (knowledge transfer), and sports prescription alerts (16). This system can not only accurately guide and improve the physical health level of college students but also bring about the integration of teaching in and out of class.

Infrastructure: The supporting environment for the intelligent governance system of college students' physical health

The collection, storage, processing, organization, analysis, and application of big data depend on the support of a stable infrastructure framework. The infrastructure provides a guaranteed platform for college students' physical health and intelligent life for big data-driven research. Its main structure is composed of hardware equipment and information technology (see Figure 2). College students' physical health intelligent service facilities cover radiofrequency identification technology, monitors, sensors, smartphones or sports bracelets, and other sensing devices as well as various information technology facilities, including memory, computing equipment, network equipment, and so on. The complete layout of a big data analysis sensor identification system has the functions of in-depth understanding, detection, and capture of sports information data, operation status of stadiums and gymnasiums, and students' needs. It is a physique and fitness big data analysis system that can comprehensively perceive, intelligently screen, and upload in real-time. At the same time, hardware equipment with high performance, high bandwidth, and a large cache is an important material premise for the analysis, storage, analysis, and application of college students' physical health big data. It should offer a solid hardware foundation for solving new challenges, such as the diversification of data sensing ports, the scale of big data analysis resource centers, and the distribution of college students' physical health intelligent services. To realize big data analysis-driven college students' physical health and intelligent services, we must also combine new technologies, such as cloud computing technology, mobile Internet, the Internet of Things, artificial intelligence, and deep learning, which is also the key basis of big data analysis of Chinese college students' physical health. We can complete big data statistical analysis, visual data analysis, semantic analysis, and predictive data analysis and find the implied knowledge relationship and other basic functions in a large number of data analyses.

Data resources and processing technology: Data mining for intelligent governance system of college students' physical health

The data resources and management stage is responsible for the organization, management, and data analysis and mining of college students' physical health big data analysis, which is mainly composed of college students' physical health big data resources and big data processing technology (see Figure 3). College students' health big data is a mass of data formed during the process of students' daily physical exercise

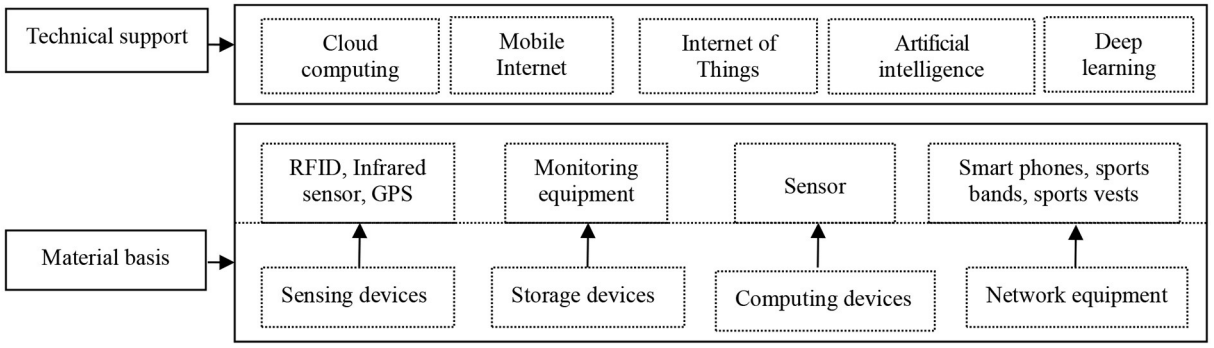


FIGURE 2
Infrastructure architecture.

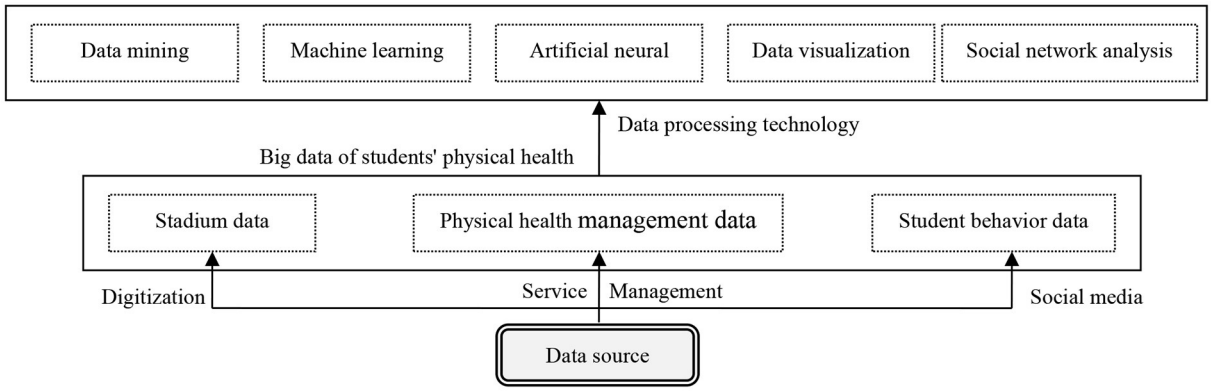


FIGURE 3
Data resources and processing architecture.

life and health management services. Generally, according to different data-formation methods, it is roughly divided into the following three types. The first is stadium database data, including venue reservation, sports knowledge base, online learning data, etc. The second is management and service data analysis. The data analysis generated during college students' daily sports includes resource construction status information records and radiofrequency identification system data analysis, access control information, monitoring information, and other Internet of Things end-user data analyses. In addition, it also involves business data analyses such as management consultation information records, training and guidance service information records, service quality evaluation feedback, and so on. The third is students' sports behavior data—that is, the behavior data left by students on the social platform, including evaluations, text, photos, videos, and so on. However, the volume of college students' physical health data is so huge, massive, and heterogeneous that new computing algorithms and protocols, more computing power, and new strategies are needed to properly and effectively combine and integrate data (17). The systematic construction of intelligent service capability is also inseparable from the support of relevant science and technology. The big data analysis of college students' physical health includes the data analysis of complex categories, such as text, pictures, sound, and videos, which requires the processing and exploration of international leading data mining technology. For example, text mining, natural language processing technology, and cognitive graphics are used to analyze text data; graphics recognition, speech recognition, video data analysis, and other means are used to analyze text, sound, and video big data; and data mining, social Internet, in-depth learning data analysis, and other means are used to analyze social media data (18, 19). To sum up, college students' physical health problems should be effectively managed and controlled through the flexible use of big data analysis, visual research, information mining, machine learning, artificial neural network, social Internet, and other analysis technologies so as to achieve the goal of maximizing the value of college students' physical health big data.

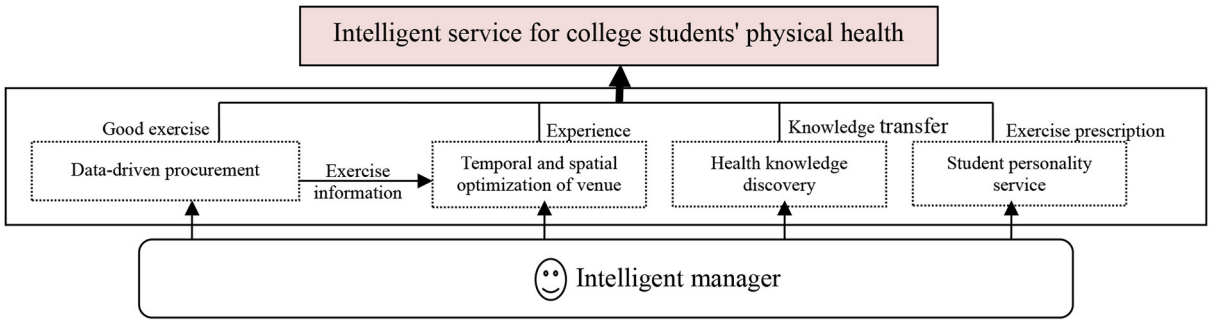


FIGURE 4
Service application architecture.

Intelligent service application: Value-shaping of intelligent governance system of college students' physical health

The intelligent service application stage is the top-level interactive port of the whole system, which directly interfaces with students and is composed of three elements: intelligent managers, students, and intelligent governance (see Figure 4). First, university intelligent managers use big data analysis methods, such as machine learning, data mining, image visualization, and artificial intelligence technology, to discover and process knowledge of college students' physical health big data and to impart professional knowledge to students. Second, college students will use the physical health intelligent management system to carry out higher-level physical exercise activities, and the formed data information will continue to enrich data resources to promote the vigorous development of college students' physical health management. In the ecological chain of intelligent governance of college students' physical health, intelligent administrators shoulder the mission of information decomposers and information disseminators, while students hold roles as both consumers and producers of information. They cooperate and coordinate with each other to jointly construct an important human resource element that is indispensable in the intelligent governance of college students' physical health. In addition, intelligent managers also mainly provide students with intelligent management of physical health of four types, including big data-driven procurement, venue space optimization, fitness knowledge discovery, and personalized services. Data-driven procurement services provide direct data resources for venue space optimization and health knowledge discovery. Health knowledge discovery and personalized services focus on the mining of physical health data and student sports behavior activity data, respectively, combining the in-depth exploration of professional knowledge and skills with the customized recommendation of professional knowledge and skills and continuing to promote the innovation of governance models, such as more accurate college students' physical health services and reference consultations, under the environment of big data and the Internet.

Big data-driven intelligent governance strategy for college students' physical health

With the comprehensive development of physical health intelligent big data services, the wide application of the Internet of Things and artificial intelligence technology, the collaborative integration of multi-source heterogeneous information, and the gradual formation of intelligent sports atmosphere, a new service strategy has been put forward for the development of physical health intelligent service systems driven by big data. This paper applies the new generation of information technology, such as big data, the Internet of Things, and intelligent equipment, to improve the physical health of college students and suggests governance strategies, such as building an intelligent governance ecosystem of school sports, establishing a support system of school stadiums and gymnasiums, and building an integrated intelligent governance model of physical health.

Top-level design: Build an intelligent governance ecosystem of school physical education

The main application purpose of intelligent technology is to help establish a physical health intelligent environment that can mobilize people's creativity by coordinating the interaction between science and technology, resources, and people. While stimulating the physical health intelligent service ability driven by students' needs, it can attract potential students and promote the intelligent development and innovation of school physical education. A physical

health intelligent management system can integrate students' physical health status into the construction of the campus sports network, resulting in a complex and integrated health ecosystem. In addition, the physical health intelligent service integrates big data analysis in the fields of physical education teaching, sports training, sports skills, after-school exercise, sports activities, academic research, and fitness management It integrates and links various sports environments, forms a network platform supporting health data analysis and decision-making, and helps the development and innovation of school sports. It also promotes the innovation of the campus sports network, and through this attracts more students to participate in the ranks of sharing sports and sports innovation. It encourages users to activate intelligent sports activities and intelligent sports services in a wider region and makes the physical health intelligent governance system truly become a network platform connecting cross-domain research, development, innovation, and various sports knowledge-intensive applications to obtain sustainable growth of new intelligent data in the campus sports network.

Interconnection of all things: Constructing the supporting system of school stadiums and gymnasiums with intelligent perception

The Internet of Things is a network of communication between people and objects and between objects, which is a perception technology for intelligent stadiums and gymnasiums. By applying high and new technologies like big data technology, Internet of Things technology, and artificial intelligence to the design and construction of campus stadiums and gymnasiums, the services of campus stadiums and gymnasiums will become more automatic and intelligent. By taking advantage of the advanced technology and equipment advantages of intelligent sensing systems and cognitive equipment, virtual reality, and intelligent robots in the Internet of Things, we can further enhance personalized, convenient, and efficient physical health services so as to make campus stadiums truly automated. For example, in terms of intelligent sensing systems, campus stadiums and gymnasiums can promote the convenience of venue resources and space management through the effective use of Internet of Things sensors. First, the use of a radiofrequency identification system, sports software, indoor navigation robots, and other advanced technical means can facilitate site reservation and management so that students can easily access site materials. Second, connections and communication among people, machines, and mobile networks can be realized through machine-to-machine technology. The module that can monitor the operation parameters of a distribution network can be installed in the power equipment of stadiums and gymnasiums to realize the real-time monitoring, control, management, and maintenance of distribution systems. Third, face recognition, infrared sensors, cameras, and other technical means help to understand students' sports behavior; infer students' needs through students' sports behavior data; and provide data support for students' personalized sports suggestions and services. In addition, through augmented reality, virtual reality, and other technical means, students can also have virtual and intelligent space service experiences of virtual and real integration (20).

Action practice: Build an integrated intelligent governance model of physical health

Through the Internet, mobile Internet, big data, artificial intelligence, and other new-generation information technologies and knowledge, we can carry out intelligent physical education and training, after-school physical exercise, physical health monitoring, and evaluation, and campus sports events. We can form an integrated physical health intelligent governance model that can accurately guide, manage, supervise, and improve the physical health of college students. Intelligent physical education is mainly embodied in information-based physical education teaching design, mobile education, and ubiquitous learning. Through big data, we can build physical education courses and online, offline, and hybrid physical education courses; independently design and develop school-based physical education curriculum resource databases, design learning methods, teaching evaluations, and other modules. It can help realize resource appreciation through the screening, evaluation, reorganization, and integration of curriculum experiences, which can meet students' online physical education learning at any time. Through big data technology, we can also carry out the integrated application of mobile teaching platforms, smart terminal teaching, and university sports services. It can be used to develop mobile teaching tools of wearable devices and university sports public service products based on big data analysis to carry out in-depth teaching research, technical innovation, system incubation, detection and evaluation, application demonstration, and other physical health intelligent services. In terms of students' intelligent after-school physical exercise and physical health monitoring and evaluation, we can use big data, the Internet, and other analysis technologies to comprehensively and dynamically manage students' physical and mental health, mastery of sports skills, and extracurricular physical exercise. The administrator transmits the sports load, physical condition, motion trajectory, and other data collected as front-end data to teachers and students through sensor devices such as smartphones, sports bracelets, and sports vests while simultaneous background calculations of the server and the global positioning system are ongoing. Teachers and systems

set the amount of exercise and monitor the situation of extracurricular exercise according to the specific situation and measurement data, then formulate exercise schedules and provide accurate guidance to students so as to reduce sports injury, realize scientific exercise and learning, and accurately serve and manage students' physical health., The new generation of information technology may be used to build the school competition platform, and an information system integrating competition qualification examinations, registration data, competition arrangements, scoring systems, and score information. Competition data analysis can be established within the platform. With this approach, both teachers and students can view the competition results in real-time; the competition situation can be recorded, analyzed, and managed. This approach allows managers, teachers, and students to master students' sports level and physical health in real-time, and data support for dynamic adjustment of physical health intelligent governance strategies can be provided. The college and students can be urged to meet relevant indicators to form an effective intelligent and dynamic integrated physical health governance system.

Conclusions

This study, based on the realistic background of big data, the Internet, mobile Internet, and artificial intelligence, sought an angle to explore the precise path and methods for improving the physical health of university students, which is a relatively new research field. The research method of traditional improvements to physical health and their innovation is of great significance to promote college students' physical health levels. The intelligent governance system of physical health driven by big data can quickly and comprehensively understand the dynamic state of the physical health of college students, while at the whole-time monitoring sports, disease health management, and dynamic adjustment of physical exercise content. Combined with a physical health management system to establish the linkage mechanism, the government, school, society, and family can use emerging information technology to master students' sports, physical development, and psychological situation; personalize push exercise prescriptions; realize students' aptitude, and precisely promote the physical health of college students to solve the problem of Chinese college students' declining physical health (21). For example, relying on cloud platforms such as the Lejian Sports app, the WeChat public platform, QQ, and online live broadcasts, the University of Electronic Science and Technology built an integrated interactive intelligent physical health management system that combines online plus offline and in-class plus after-class data. Club teaching or online teaching and practice requirements are mainly implemented for students in class, and offline guidance and cloud interactive practice are implemented for students after class. Analysis of the sports data platform revealed that, since the construction of intelligent governance systems for physical health, the daily number of extracurricular sports has increased greatly, the qualified rate of physical health has improved year by year, and the physical quality level has also improved significantly.

It is suggested that schools at all levels should make full use of modern information science and technology in the future, monitor the physical health of college students, monitor their physical activities, set up a physical activity guidance and service platform, guide students to exercise scientifically, and strive to improve the physical health level of students. How to make full use of big data, artificial intelligence, intelligent equipment, and the Internet of Things as part of a new generation of information technology to improve students' physique health as well as to create intelligent physical health service models for precise improvements in students physical health patterns remain future research directions to be explored.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CD and QY studied the current situation of college students' physical health and its informatization governance problems, the construction of governance system, and intelligent governance strategies. GL, ZZ, and YL participated in the exchange and discussion. CD drafted and revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Basic Scientific Research Operating Expenses Project of the Central

University of China (Grant No. ZYGX2019J143), the Educational Research Project of Sichuan Province (Grant No. SCJG20A169), the Ideological and Political Demonstration Course Construction Project of UESTC (Grant No. 2021KCSZ0196), and the Relay Action Plan for Young Teachers in Philosophy and Social Sciences of UESTC in 2021.

Acknowledgments

Thanks to the authors for their efforts in this paper, especially to the authors of the works we cite or refer to in the paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Tong EB, Cheng ST, Feng W, Wang YS. Dilemma and solution: research on the integrated development of sports big data industry. *J Xi'an Phys Educ Univ.* (2021) 38:434–40. doi: 10.16063/j.cnki.issn1001-747x.2021.04.007

2. Dong YQ, Zhong JW, Ding F. The development path of China's sports industry in the era of big data. *Sports Cult Guide.* (2018) 12:76–81. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=TYWS201812017&uniplatform=NZKPT&v=4QtTPa5y5uPWIjD9tWjLFYRMXEBLP5t5a-qMRORC7-t0_WQRjthEt9zLpBgGtB5C

3. Yu WW. Financial risk control of sports industry in the era of big data. *Commun Finance Acc.* (2018) 32:116–8. doi: 10.16144/j.cnki.issn1002-8072.2018.32.027

4. Xie JL, Sun JH, Cao L. Opportunities, challenges and countermeasures of sports industry development in China in the era of big data. *J Shanghai Univ Sport.* (2015) 39:59–63. doi: 10.16099/j.cnki.jsus.2015.04.012

5. Zhang LJ, Sun YP. Sports precision teaching model driven by big data. *J Tianjin Univ Sport.* (2022) 37:174–80. doi: 10.13297/j.cnki.issn1005-0000.2022.02.008

6. Dou L, Chen HW, Qian C. Study on the value and model of intelligent physical education class in colleges and universities. *Sports Cult Guide.* (2018) 11:136–46. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2018&filename=TYWS201811028&uniplatform=NZKPT&v=4QtTPa5y5uNhfs9nWCWiFRFpkYrslC3QdlwT62h1EskUDLXv98rLcqqcHKX435wX

7. Zhang Q, Wang JH. Research on operation and management of smart sports venues in China in the new era. *J Wuhan Inst Phys Educ.* (2021) 55:62–9. doi: 10.15930/j.cnki.wtxb.2021.11.009

8. Liu Y. Research on the construction and promotion strategy of smart sports service system for the elderly in Chinese community. *Sports Sci.* (2021) 41:51–64. doi: 10.16469/j.css.202101006

9. Song YX, Liang W. Challenges and governance of personal information protection in the era of smart sports: A case study of public smart sports facilities. *Sports Sci.* (2021) 41:51–9. doi: 10.16469/j.css.202110007

10. Fan M, Zhang XL. Value implication and practical enlightenment of the study of physical education from the perspective of curriculum ideology and politics. *J Shenyang Sport Univ.* (2022) 41:67–75. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2022&filename=SYTB202201009&uniplatform=NZKPT&v=UrDC0OIhDqeAJXwQ57NDnqO-BkfltmQbIJjwkM8eNSluXXkT6JVr8yxCoMTaBmCh

11. Yin C, Li Y. The influence of family education on students' physical exercise habit. *Sci Educ Guide.* (2013) 9:79–82. Available online at: http://zlf-cqvip-com.vpn.uestc.edu.cn:8118/article/detail.aspx?id=48506485&from=zk_search

12. Chen BL, Lu YZ. The influence of family on sports consciousness and behavior of college students. *Sports Cult Guide.* (2008) 1:100–2. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2008&filename=TYWS200801034&uniplatform=NZKPT&v=8SdotHIEQXH4wnY1jJLXoo4QhkqtI3SIUje0NlfKdkdx1CJtiuvOk4SfWDVPkZzy

13. Yang DY, Yang WW, Huang DZ, Liao P. Performance, causes and countermeasures of deviations in the implementation of school physical education policy – taking the policy analysis of sunshine sports as an example. *J Wuhan Inst Phys Educ.* (2014) 48:78–82. doi: 10.15930/j.cnki.wtxb.2014.01.015

14. Xing YM. Core elements of modern higher vocational education informatization construction. *Vocat Educ Res.* (2016) 35:41–4. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2016&filename=JLKF201601010&uniplatform=NZKPT&v=457dvLtVOGww1twDs2vuuyWCj1yiXyVzKqUE6RBt3pUzjs2T5EOh9Txh-biY9PD_

15. Yu QY, Pan XF. Main content and implementation path of college sports informatization construction. *J Phys: Conf Ser.* (2020) 1575:1227–32. doi: 10.1088/1742-6596/1575/1/012184

16. Hong L, Zhou LN, Chen CQ. Research on the construction of data-driven smart information service system in libraries. *Lib Inf.* (2018) 2:8–15. Availble online at: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2018&filename=BOOK201802002&uniplatform=NZKPT&v=UbgjYyHB_QMlryfu6fvr7XuZBegvG3HfnsJ7Lg5lNdWotok7NttFYo9R3ubCr1bx

17. Sellami M, Elrayess MA, Puce L, Bragazzi NL. Molecular big data in sports sciences: state-of-art and future prospects of omics-based sports sciences. *Front Mol Biosci.* (2022) 8:815410. doi: 10.3389/fmolb.2021.815410

18. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage.* (2015) 35:137–44. doi: 10.1016/j.ijinfomgt.2014.10.007

19. Zou HB, Chen HM, Dey S. Understanding library user engagement strategies through large-scale twitter analysis. In: *1st IEEE International Conference on Big Data Computing Service and Applications (Big Data Service).* IEEE Computer Society (2015). p. 361–70. doi: 10.1109/BigDataService.2015.31

20. Li HG, Manickam A, Samuel R, Dinesh J. Automatic detection technology for sports players based on image recognition technology: the significance of big data technology in China's sports field. *Ann Oper Res.* (2022) 3:1–18. doi: 10.1007/s10479-021-04409-1

21. Yu QY, Deng CL. Research on the construction of intelligent service system of adolescent students' physical health based on big data, Abstract of the Papers of the 11th China Sports Science Conference. *Chin Sports Sci Assoc.* (2019) 6535–7. doi: 10.26914/c.cnkihy.2019.031908

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fpubh.2022.926229

Check for updates

OPEN ACCESS

EDITED BY
Celestine Iwendi,
University of Bolton, United Kingdom

REVIEWED BY
Sonali Vyas,
University of Petroleum and Energy Studies, India
Sandeep Kautish,
Lord Buddha Education Foundation (LBEF), Nepal
Rajat Bhardwaj,
Chitkara University, India

*CORRESPONDENCE
Surbhi Bhatia
sbhatia@kfu.edu.sa

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 22 April 2022
ACCEPTED 27 June 2022
PUBLISHED 10 August 2022

CITATION
Bhatia S, Alojail M, Sengan S and Dadheech P (2022) An efficient modular framework for automatic LIONC classification of MedIMG using unified medical language.
Front. Public Health 10:926229.
doi: 10.3389/fpubh.2022.926229

COPYRIGHT
© 2022 Bhatia, Alojail, Sengan and Dadheech. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An efficient modular framework for automatic LIONC classification of MedIMG using unified medical language

Surbhi Bhatia[1]*, Mohammed Alojail[1], Sudhakar Sengan[2] and Pankaj Dadheech[3]

[1]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al Hasa, Saudi Arabia, [2]Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, India, [3]Department of Computer Science and Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan (SKIT), Jaipur, India

Handwritten prescriptions and radiological reports: doctors use handwritten prescriptions and radiological reports to give drugs to patients who have illnesses, injuries, or other problems. Clinical text data, like physician prescription visuals and radiology reports, should be labelled with specific information such as disease type, features, and anatomical location for more effective use. The semantic annotation of vast collections of biological and biomedical texts, like scientific papers, medical reports, and general practitioner observations, has lately been examined by doctors and scientists. By identifying and disambiguating references to biomedical concepts in texts, medical semantics annotators could generate such annotations automatically. For Medical Images (MedIMG), we provide a methodology for learning an effective holistic representation (handwritten word pictures as well as radiology reports). Deep Learning (DL) methods have recently gained much interest for their capacity to achieve expert-level accuracy in automated MedIMG analysis. We discovered that tasks requiring significant responsive fields are ideal for downscaled input images that are qualitatively verified by examining functional, responsive areas and class activating maps for training models. This article focuses on the following contributions: (a) Information Extraction from Narrative MedImages, (b) Automatic categorisation on image resolution with an impact on MedIMG, and (c) Hybrid Model to Predictions of Named Entity Recognition utilising RNN + LSTM + GRM that perform admirably in every trainee for every input purpose. At the same time, supplying understandable scale weight implies that such multi-scale structures are also crucial for extracting information from high-resolution MedIMG. A portion of the reports (30%) are manually evaluated by trained physicians, while the rest were automatically categorised using deep supervised training models based on attention mechanisms and supplied with test reports. MetaMapLite proved recall and precision, but also an F1-score equivalent for primary biomedicine text search techniques and medical text examination on many databases of MedIMG. In addition to implementing as well as getting the requirements for

MedIMG, the article explores the quality of medical data by using DL techniques for reaching large-scale labelled clinical data and also the significance of their real-time efforts in the biomedical study that have played an instrumental role in its extramural diffusion and global appeal.

KEYWORDS

MedIMG, deep learning, LIONC, accuracy, natural language processing

Introduction

The digitization of documents has created plenty of new opportunities for liberalising data collections for NLP pipelines like machine translation as well as web search engines. Automatically creating captions for a certain image is problematic, while the foundation of understanding the scene is partly due to computer vision's exceptional achievement. Books, manuscripts, letters, bills, catalogues, and other digital documents (MedIMG) are machine-printed, handwritten, or radiology images (1). Learning effective representations of document images that are rich in preserving both lexical as well as semantic information is the only way to gain access to content-level in such a vast numerical corpus. In this paper, we focus on encoding the word MedIMG in a featuring space that conserves its verbal data, and we limit our stability by only using the number of declarations. The restricted availability of large-scale labelled datasets within the medical realm justifies their construction, as they play a vital role in developing medicine using Machine Learning (ML) techniques (2). Clinical text data like physician notes as well as radiology reports must be annotated using labels that include specific information, namely the kind, features, and anatomical area of disease, for more effective use. However, since text contains highly specialised language, ambiguous terms, and narrative sentences, labelling text is challenging and must be done manually by health experts. Some of the things that make it hard for clinical text to be used in considerable medical research and mostly with DL approaches is that it does not have annotations (3).

This study can be presented as a completely innovative descriptor of word pictures that may be utilised to construct MedIMG-based applications (X-rays, handwritten and printed documents, radiology images) (4). To obtain the UMLS concept, annotations/captions use picture data, which will subsequently be compared to the images' original text captions. The images utilised are clinically relevant radiological images, and the captions were written by medical professionals. Moreover, DL approaches are used to extract medical knowledge using clinical documents. We believe that the classification of scanned documents into words is provided to us either through underlying data or through an external word suggestion approach, which may be chaotic. In this paper, a word-level description is presented as a model. It has to do with extracting information from medical records that include radiographs. In addition, MetaMapLite pays attention to real-time processor speed and then begins with a restricted range of essential functionalities like the longest-term match as well as negation detection. MetaMapLite is presented in this work, and its execution equals that of the existing Prologue operation of MetaMap is assessed (5). We examine whether employing high-resolution MedIMG rather than downscaled and low-resolution MedIMG can lead to improved clinical work performance. We anticipate that increasing the size of an image would improve MedIMG segmentation's work performance (6). Throughout all input resolutions, the experimental results show how to stack groups across determinations beat each discrete learner. We additionally give easily understandable Ensemble Model (EM) scale weights, implying that multi-scale features are required for automated detection of a higher level of MedIMG. Extensive research is being conducted in the field of handwritten character recognition. Handwritten character recognition has been developed by many users. A few systems have been examined: Using fuzzy logic, a character recognition system was created (7). They have designed a system that is compatible with VLSI technology. Their system for recognising characters can't be changed or messed up. They have incorporated hammering NN into their system. It has been found that NN can be used to recognise handwritten characters in a novel way. A Kohonen Self-Organising Map (SOM), an unsupervised NN, has been utilised.

The following contributions are focused on this paper:

- Information Extraction from Narrative Clinical Texts: An automatically generated mapping system connects basic phrases as well as codes collected throughout several clinical terminology and programming systems for Systematised Terminology of Medicine-Clinical Terminology concepts to standardise multi-site heterogeneous Electronic Healthcare Record (HER).
- Automatic Classification for MedIMG with Image Resolution: Next, an innovative large-scale dataset of X-rays is used to accurately assess its chances of attaining increased clinical work performance using high-resolution visuals rather than downscaled low-resolution MedIMG.

MedIMG, as well as findings of multi-label annotations, are being developed.

- Hybrid Model to Predictions of Named Entity Recognition (NER)by RNN + LSTM + GRM: Our goal is to present an innovative UMLS-based RNN+LSTM+GRM for NER tool and compare its outcome to state-of-the-art, broadly utilised, accessible public medical NER tools.

Apart from trying to implement and obtain results for clinical testing and MedIMG, this article refers to the study of health data, the innovation of using DL methods for obtaining large-scale labelled MedIMG datasets (8), as well as the relevance of its applications in the medical research that has played a role towards its extramural diffusion and global reach. This paper summarises the journey in such a way that the reader is guided through multiple disciplines, from engineering to medicine, all the way up to the ethical implications of implementing Artificial Intelligence (AI) in treatment.

Literature review

The health care industry must have commonly recognised and advertised the partial use of EHR systems throughout the previous decade. The Health Information Technology to Economic and Clinical Health (HITECH) Act of 2009 (9) motivated medical hospitals and clinics to implement EHR systems. Though HER systems are planned for operational purposes, they were later used to process data. All of the information recorded by the EHR system could now be analysed. The enormous data acquired by progress notes, combined with the rapid adaptation of EHR systems, has resulted in a significant research area of predictive medical analytics, which employs narrative progress notes. For many years, conventional Machine Learning (ML) models were used to perform predictive analysis, mainly in the medical domain. Because of the operational excellence of DL methods in recent times, many have been widely used in biomedical disease prediction. For example, using EHR data, researchers created a Long Short-Term Memory (LSTM) network model for predicting heart failure. A Deep Convolutional Neural Network (CNN) is used to predict diabetes. In addition, a CNN methodology to detect colorectal cancer, which uses diagnoses as well as medication of patients in the HER, was developed (9). Using real-world Pashto/Urdu text, another research was performed, and this text can be rotated in any direction, not just in a straight line. The results were analysed using various ML techniques, including Scale Invariant Feature Transformation (SIFT), LSTM, and Hidden Markov Model (HMM) to detect rotated text. In this study, the LSTM got 98.9% accuracy, while HMM-based techniques got 89.9% accuracy and SIFT got 94.3% accuracy.

Multiple successful efforts were made in the earlier days, focusing on various factors like data modality (radiology and handwritten images), nature of depiction (firmed/adjustable length), and also by categorising of entrenching structure. It is divided the debate of associated studies are: (a) traditional techniques developed by means of flexible length depiction plans, (b) static length depiction attained by means of a bag of words structure, and (c) different classifier models with learned representation implemented on a maximum of hand-crafted features as well as in DL networks (10).

A data analysis Language Model (LM's) goal is to approximate the probabilistic model on classifications of distinct entities or tokens. This work includes a variety of Natural Language Processing tasks (NLP), and the majority of NLP tasks necessitate the use of NLP models. An NLP model shown in such a task describes the possibility distribution function of word sequences and stemmed word-token sequences (11). This approach's efficacy was tested using a variety of datasets, including three well-known Handwritten Text Recognition (HTR) datasets, as well as three cutting-edge Optical Models for text recognition and eight different spellcheck correction techniques, including both traditional statistics and current approaches utilising NN in NLP. They made a model for spelling correction that was tested statistically with HTR system metrics. In the tests, the average sentence correction rate was 54% higher than with the most advanced decoding method. An LM, on the other hand, can define probability distribution function on sequences for some other type of token, like bytes/characters. So, this is the issue of LM character-level, which models language to be distribution functions on sequences to characters rather than words- for example (12), without a clear and specific notion of sentence-level boundaries. It has been demonstrated that for learning to forecast the next character based on earlier characters, such models also can understand inner depictions, which obtain syntactic and semantic properties and then produce accurate grammar transcripts with words, subclauses, quotations, and sentences.

Word Embeddings (WE) (13), recognised as word vectors and distributed representations, are complex depictions of words in a low-dimensional actual vector space that overcome the dimensionality blow of n-grams. Neural Word Embeddings are those WE studied by a Neural Network (NN). In terms of computer visualisation, these embeddings are analogous to image embeddings acquired as images in CNN models' hidden layers.

Convolutional neural network has been used to process the data with a grid-like structure and spatial patterns. CNNs have been effectively applied to various forms of data wherever they can be presented as discrete units grouped in periodic intervals together in a grid-like variety of dimensions. The local spatial features of images are determined by CNN and reported to generate relatively high features. The first CNN layer generates feature maps that include spatial features that were extracted locally. This shows how high-level features and local spatial features are made. Despite the fact that they were initially

developed for picture data, it is presented as a 2-D set of pixels. Data of Time-series, which is presented like the 1-D grid with examples collected during periodic intervals, then text, which can be thought of like a 1-D grid of ordered tokens, are two examples (13).

Recurrent Neural Networks (RNN) are developed to evaluate sequential data by recursively sharing the same weights across successive time steps "t." At each time step, a member of an input sequence is therefore fed into the network. RNNs can be designed in various ways based on recurrent connexions within their units. LSTM (14) is a sort of gated RNN which uses self-loops to create routes with long gradient flows and has been found to learn long-term dependency easier than vanilla RNN designs. Gated RNNs provide temporal pathways with a derivative that does not explode or vanish, similar to leaky units. These pathways include connexion weights that can vary for every time step, enabling knowledge to be accumulated or forgotten by erasing the previous state.

A Multi-Label Text Classification (MLTC) (15) task is one that involves extracting many medical categories from radiological data. MLTC is indeed an NLP task that allows you to apply one or even more labels per document. This differs significantly from the typical binary but rather multi-class classification issues that have been extensively researched inside this machine learning literature (16). Binary classifier treats classes as separate target variables, which is naturally inefficient for MLTC because label dependencies cannot be exploited. A mutually exclusive assumption concerning classes is used by multi-class classifiers. According to this concept, each document should have one class, which is not valid in multi-label setups. This section's related works are primarily concerned with MLTC (16). A novel method for recognising handwritten characters has been advanced using NN. They used an unsupervised NN called the Kohonen Self-Organising Map (SOM). They developed a system that can be used for character recognition as well as the recognition of other Indic languages. If the handwritten characters are not effectively segmented, the system occasionally provides false results. One of the authors can use the handwriting of a human to verify their authenticity. In their system, the author has implemented an FFNN with multiple layers. One sophisticated hypothesis in this paper is that the handwritten alphabet's height and width are unique to the writer's individual hand size and handwriting. It has been demonstrated that a patient's handwriting can be used to identify them. Hyper-parameters and model parameters can differ significantly, as demonstrated by (7). Some hyper-parameters, such as the number of epochs, hidden units, hidden layers, Learning Rate (LR), kernel size, and activation function, must be set before training begins. It is stated that if the hyper-parameters are selected improperly, CNN performance may suffer. Some CNN models have 27, 57, 78, and 150 hyper-parameters, respectively, for AlexNet, VGG-16, GoogleNet, and ResNet-52. In the field of handwriting recognition, implementing researchers play a big role in making sure that CNN parameters are set up well so that recognition performance can be improved.

We present an automated picture denoising method based on quasi-periodic denoising. The Fourier transformation is used to detect the image's high amplitude noisy spectrum. The proposed model for noise filtration has a threshold value of 30. The threshold value is updated in response to user input. Thresholding, as well as filtration-based models, are used to minimise the noisy components (17). Finally, an adaptable restoration filter is used to obtain the noise profile's exact outline. A method for denoising images using a quicker and more flexible CNN is provided. This study takes into account various levels of noise that use a non-uniform noise level map to eliminate the spatial noise. For applying local low-priority matrix recovery with global spatial-spectral variation, a hyperspectral image denoising algorithm is suggested. A unified mixed Gaussian noise plus sparse noise removal model is recommended in this research article. To separate the filtered image regions from sparse noise, the hyperspectral images are first segmented into overlapping local patches, and thus the matrix recovery scheme is used. This work rebuilds the image with global acceptance (18). As a result of this, a machine is trained to correctly identify Pashto handwritten characters that have never been seen before. With the help of this dataset (7), we trained and tested three FFNN models (Model 1 with a 1-ReLU layer, Model 2 with 2-ReLU layers, and Model 3 with 3-ReLU layers) using the backpropagation algorithm. Sample images of user-defined handwritten text were used to test the proposed overall system performance. This preprocesses images to prepare them for CNN training (19). The input document is then segmented by line, word, and character. Multiple experiments yielded results deserving of high recognition. At one point in the character segmentation method, they were able to get up to 86% accuracy.

The labelled datasets of the most frequent named entity recognition tasks are used widely and are publicly available. CoNLL-03, a shared study from the Conference on Computational NLP Learning in 2003, the MUC-7 dataset (20), and also the Onto Notes corpus consists of around ten named entity types as well as seven various entity forms like TIME and DATE, which are the three most common corpora used to train such extractors. The jobs are typically modelled using BIO upgraded labels–separate labels for the beginning, inside, and outside tokens of an entity.

Mathematical model

Generation of MedIMG caption

Encoder of convolutional features

Our model provides a caption X that is encrypted as a series of 1-n encrypted words from a single input image Eq. (1).

$$X = \{X_1, ...X_C\}, X_i \equiv R^n \quad (1)$$

WHERE the size of the syntax by K and the length of the tagline by C.

To retrieve a sequence of a feature vector that refers to an annotation vector (21), we employ a CNN. The extractor generates Y (vectors); each one is the d-dimensional description of image segmentation, Eq. (2)

$$Y = \{Y_1, ...Y_L\}, Y_i \equiv R^d \tag{2}$$

In contrast to prior work, which used a fully connected layer, we extract the features from such a lower convolutional layer to derive a correspondence between feature vectors and sections of the 2-D image. By weighting a subclass of whole feature vectors, the interpreter can preferentially focus on some areas of a picture.

Decoder of LSTM

We apply an LSTM network to construct captions by producing one word at a time based on the context vector, the initial Hidden State (HS) (22), and earlier generated comments, Eq. (3)

$$\begin{aligned} a_t &= (P_i Q_{t-1} + R_t T_{t-1} + S_i s_t + b_i) \\ b_t &= (P_i Q_{t-1} + R_t T_{t-1} + S_i s_t + b_f) \\ c_t &= f_t c_{t-1} + i_t + \tanh(P_c Q_{t-1} + R_c S_{t-1}) \\ d_t &= (P_i Q_{t-1} + R_{t-1} + b_i) \\ e_t &= O_t + \tanh(c_t) \end{aligned} \tag{3}$$

The input, output, forget, HS, and memory of the LSTM are represented by at, bt, ct, dt, and et, respectively. P●, Q●, R●, and S● are weight matrices as well as biases that have been learned. An embedding matrix is PϵQ mxn. Let "P" and "Q" stand for LSTM and embedding dimensions as the logistical sigmoid initiation.

Figure 1 shows an LSTM cell, along with bolded squares indicating projections of the learned weight vector. Every cell learns how to balance its input components while also studying how to control its memory supply. Also, it learns weights that delete the memory cell as well as weights that govern the emission of memory. The Context Vector (CV) is a dynamic depiction of a related component of such visual input to a given point in time (t). By defining a method, "ϕ," which calculates CV beginning Annotation Vectors (AV) ai, I = 1,..., loc corresponding to such features obtained at various picture settings. The mechanism generates a positive weight I for each location "loc" that can be taken as whichever is the probability of such location "loc," it remains the best place to develop and to generate the succeeding word (stochastic attention method) or else the virtual importance to bounce to location "loc" in merging the αi's (αi's blending mechanism) (23). Non-linear and linear data can be learned by NN. Where the neurons are thought to consist of multiple layers. An MLP is a directed graph of input nodes that connects the input and output layers. As a supervised learning technique, backpropagation is used to train the NN in this case. An attention model, for which we apply a multilayer perceptron conditioned here on preceding HSt−1, computes the weight αi of all vector annotation AI. The HS changes to be the output of RNN that progresses through their output sequence: "here" the next network search is determined by its previous series of words (Figure 1).

$$\begin{aligned} HS_t &= f_{ATTN}(ai, HS_{t-1}) \\ \alpha_t &= \frac{Exp(HS_t)}{\sum_{n-1}^{Loc} Exp(HS_i)} \end{aligned} \tag{4}$$

From Eq. (4), the CV (St) is calculated after the weights have been computed, while "ϕ" is stated as a function, which gives a specific vector based on a sequence of AV as well as its related weights. An average of such AV fed *via* two distinct MLPs (Intc, Inth) predict the LSTM's initial memory state as well as HS, Eqs (5), and (6)

$$S_t = \alpha[a_i], [\partial_i] \tag{5}$$

$$S_0 = F_{\text{int}}, a(\frac{1}{loc}\sum_{n}^{loc} a_i), h_0 = F_{\text{int}}, h(\frac{1}{loc}\sum_{n}^{lco} a_i) \tag{6}$$

An average of AV fed into two independent MLPs predicts the LSTM's baseline memory state and not the HS (Intc, Inth). For calculating the output word probability in this study, we are using a deep output layer. Cueing out from the image, the predefined word, as well as the Decoder State are all used as input (DSt), Eq. (7)

$$P(Y_t|a, Y_1^{t-1}) \equiv Exp(loc_0(E_{t-1} + L_t DS_t + L_t R_t) \tag{7}$$

Named entity recognition using LONIC

We designed a rule-based approach to automatically normalise a local testing name towards a Logical Observation Identifiers Names and Codes (LONIC) code of MedIMG utilising data from several medical systems. Following tokenization of the input lab testing names, specific items are recognised, and relevant LONIC codes get automatically mapped depending upon that coding rules (24). Figure 2 depicts an overall view of the MedIMG TestNorm system's module,

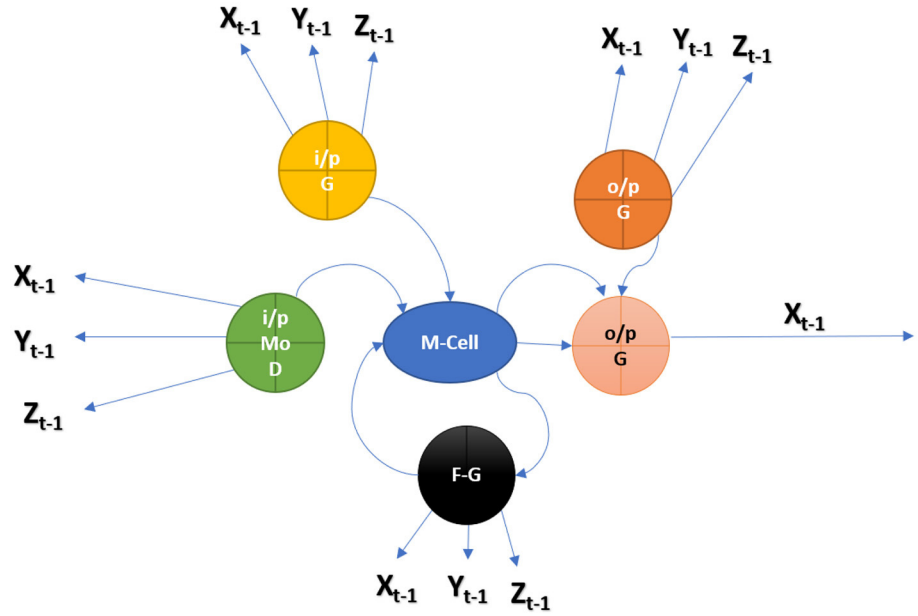


FIGURE 1
Cell format of LSTEM.

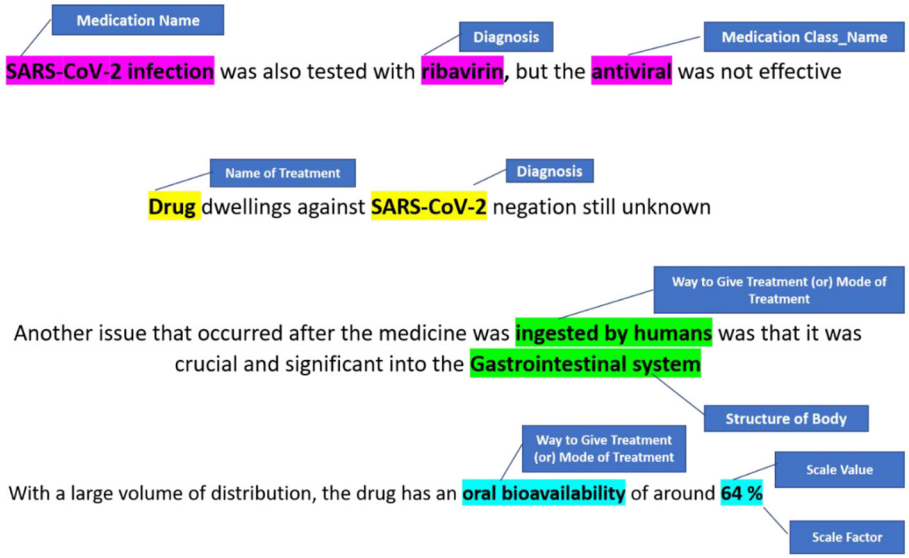


FIGURE 2
Example of NER for clinical symptom extraction.

which primarily includes NER as well as LONIC map-based modules, using inputs from lexicons/coding rules.

NER is a text processing technique that recognises words belonging to specific Named Entity categories (25). This is a crucial tool in NLP to extract information from documents. The network is trained to self-learn subject-related terminology with the aid of a dense architecture with DL models. All named entities inside the document are identified by NER. The NER annotator, in particular, uses multiple ML algorithms that can tag entities using standards for identifying numerical entities such as time/date formats. The goal of NER is to find the symptoms that

are associated with the sentences collected from EHR system (26).

Logical Observation Identifiers Names and Codes is a database with a vocabulary coding system designed to make detecting and presenting medical laboratory observations more uniform and global. Component, System, Method, Time, Property, and Scale are six fundamental axes used by LONIC to characterise each notion (27). Nine of these were included with our MED IMG entity classifications. Component, System, Method, and Quantitative/Qualitative define whether a test produces a quantitative/qualitative result, and Institution, that specifies a testing kit maker is our five root groups (28). The entity recognition process is divided into two phases: (a) a preliminary phase, which integrates dictionary lookup with regular pattern matching, then (b) a class of activities phase that translates unclear tags from preliminary step to final tags using a set of predetermined rules (Table 1).

Many data can be recorded and categorised with their appropriate category within the first stage, but some problematic words will have to be examined further. For instance, the letter "IA" could be mapped to either a "method" and is the abbreviation of "immunoassay" or a "system," which is the state of "Iowa." To select the relevant semantic classifications for such phrases, we built context-based criteria. Such investigation code must be now a LONIC code so that an identifier can be sent. A LONIC code (Figure 3) may be provided additionally towards a legible registered displaying name, which correlates for specific internal purposes to the degree permitted by the protocol. These characters can be classified with high accuracy because they have a range of data about the flow of the text being written. This makes it simpler to distinguish between different characters in the text. After the raw text has been preprocessed, a feature extraction step is performed to determine important details about each character, such as loops, quality of decision making, aspect ratios, etc. A proposed model is then used to analyse these features. Due to the manual step of extracting features and the fact that DL models can only do so much, their performance is not very good. In this section, we analysed the data to determine whether the researchers should revise certain sections of the paper. Tables listing handwritten names, layout, results obtained, and remarks are preferable because most empirical studies use the same algorithm with comparable performance datasets.

The use of LONIC across many application fields is becoming progressively acknowledged worldwide over the last 25 years (29). According to the Regenstrief Institute's LONIC webpage, there were about 85,000 clients of LONIC in over 176 countries, and LONIC is becoming the "lingua franca for medical data interchange."

Construction and structure of LONIC

Whatever the laboratory test, medical observation/clinical examination is clearly determined in Terms of six axes, as per the following LONIC concept (Figure 4): Component, Property, Time Aspect, System, Scale, and Method Type. The above six axes are given in the LONIC database by corresponding columns (30).

The technique of investigation in the sixth axis is limited in that it must be specified as required for precise determination to assess the examination under issue, and this differentiation is crucial in clinical practice. As a result, this column is frequently empty. All six axes add up to the unique, complete generic identity for the examination, as per the LONIC nomenclature guideline (31). Each form of investigation, measurement, or observance, which varies in any of these six axes, is assigned a unique LONIC code, as well as an individual entry inside the LONIC database. These LONIC codes are being posted sequentially to the access; unlike a categorisation, the codes need not indicate links within concepts. It is made up of a numeric multi-digit code of only a check digit at the end that is generated using the Mod10 method (32). The process is documented inside the LONIC User Manual primarily by Regenstrief Institute and the LONIC Board (33) (Figure 5).

Towards summation, all medical adherence/clinical assessment is represented with seven axes: the previous LONIC encryption as well as the recognised name, which consists of the six axes defined on top. Figure 5 shows the example of measurable urine sugar quantity using test strips.

The probable discrimination between purposes of the similar chemical, such as glucose, with dissimilar test materials like urine, cerebrospinal fluid, or serum, as well as through different procedures, scaled type, also attributes, occurs using the six axes (34). Only the property is clearly defined on the Property axis, not the allocated measuring unit and its representation. To accurately attain the specific LONIC code for such an evaluation issue, it is vital to recognise and describe all six axes. By taking a glucose amount, for example, that is shown by determining [Property = SCNC], a molar substance absorption would be distinct from determining a mass absorption [Property = MCNC]; collected results will obtain different LONIC codes in a similar example category. However, whether the LONIC code is represented in component g/L or mg/dl, and specific components are being stated at machine-readable procedure, it has no bearing on the code's uniqueness (35).

TABLE 1 Summary of database and LOIN code.

Database type	Meaning	Abbreviation	Full name
Characteristic	Measurement component/anion	Acnc	The surface of the hepatitis B virus
Area	The type of measurement; includes a conceptual model (dosage, composition, factor, subset, and exceptional)	P_t	Single point time
Monitoring	Identifies the testified measurement's time things, which are treated as a SPT	Ser/Pls	Plasma vs. serum
Method	It refers to a set of data/specimens being measured	Q_n	Qualitative research
Level	Qualitatively channeling's the monitoring value	IA	Biosensors based on metabolites
Procedure	It involves the procedure used to extract the standard size		

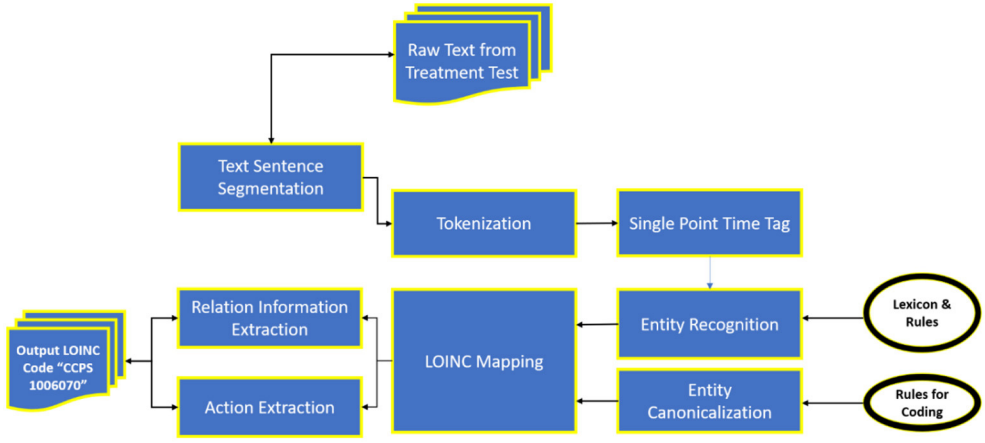


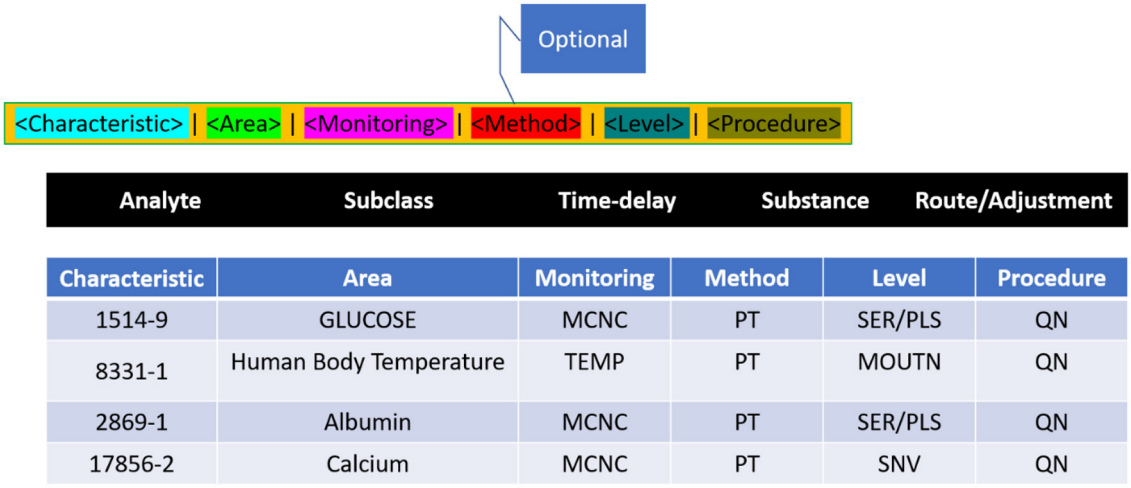
FIGURE 3
Mapping of LOIN codes and rules.

Proposed methodology

Investigation of clinical texts for information retrieval

An end-to-end "pipeline" for obtaining crucial clinical information from narrative texts is being designed. Negation detection is used to filter these features, and the residual features are mapped to SNOMED-CT terminology.

A feature extraction pipeline of clinical text is depicted in Figure 6. The TextRank algorithm (36) is used to implement the content summarising module. Primary clinical information such as problems, procedures, and tests are extracted using the CLiNER concept recognition algorithm. The features inside a negated context are subsequently filtered using an updated Negex algorithm. This part of the pipeline is a secure in-house online application that allows users to submit files that contain descriptive clinical notes and extract essential clinical features together with their contexts. From this, you can pursue one of two paths: (1) Use MetaMap to map the retrieved features towards the SNOMED-CT terminology systems and filter out unmapped features. Eliminate standard terms (e.g., "Human Body Structure," "Clinical investigations," "Biological agent") from the top two levels of the SNOMED-CT concept tree using the hierarchical system of SNOMED-CT and MetaMapLite; (2) use the terminology mapping tool



Characteristic	Area	Monitoring	Method	Level	Procedure
1514-9	GLUCOSE	MCNC	PT	SER/PLS	QN
8331-1	Human Body Temperature	TEMP	PT	MOUTN	QN
2869-1	Albumin	MCNC	PT	SER/PLS	QN
17856-2	Calcium	MCNC	PT	SNV	QN

FIGURE 4
LONIC nomenclature.

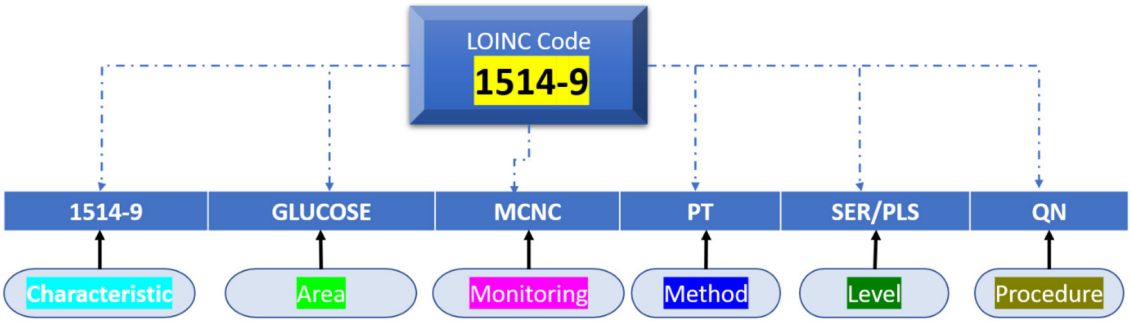


FIGURE 5
Quantitative measure of LONIC name and code.

formulated to map those particular concepts to SNOMED-CT effectively. These standardised concepts can be aggregated in input features that can be directly fed into knowledge discovery ML algorithms. We illustrate the usage of EHR information extraction strategies for several patients in this section (37).

Automatic classification of MedIMG by image resolution

The IAM Handwriting Database is applied to arrange the handwriting, as previously stated. To begin, everyone must first register on the website and retrieve a folder of photos as well as XML files containing the text location in the image. In this study, N's do not require significant pre-processing of raw data (33). As a result, it will not make any changes or adjustments in these images but rather send the entire image and just patches of text to the NN.

Generating patches of data

Instead of transferring specific sentences or words, merely random patches containing text are allowed to train the NN that comprehends the writing style of doctors. Cropped photos with a random size of 113 × 113 can be used to generate patches from each sentence. Figure 7 below is a mosaic of eight of these patches.

A simple generating function is constructed using Python to traverse over each text, allowing for cropping as well as generating random patches for images. We can limit the number of patches caused for each image to 70% of the total generated patches.

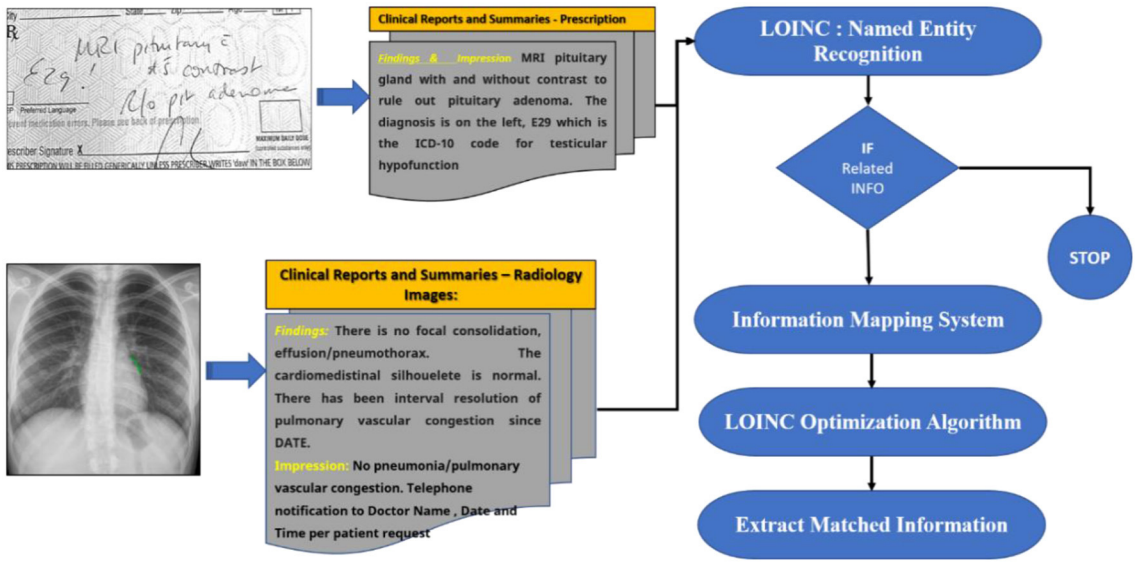


FIGURE 6
A proposed model for extracting patient treatment features from MedIMG.

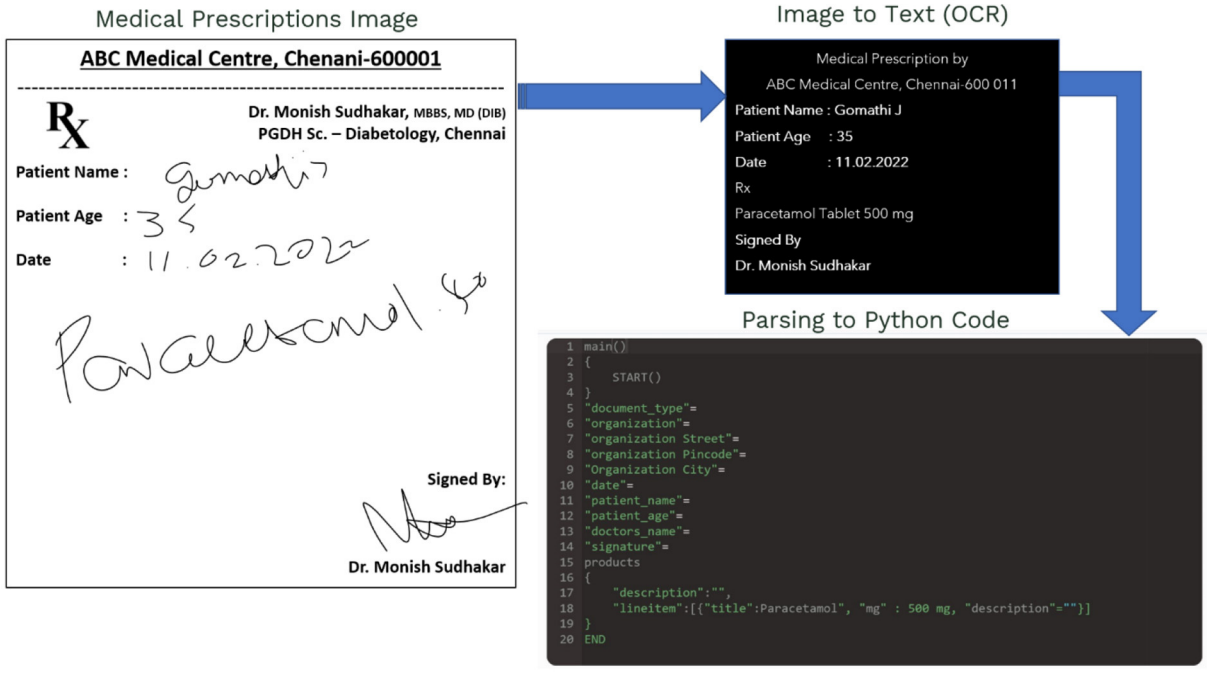


FIGURE 7
Cropping and generating random patches of MedIMG.

HWNet v2

HWNet v2 is a new version of the HWNet network. We restricted the number of convolutional layers in our basic HWNet architecture to five, which is equal to such suggested layers at AlexNet. Innovative structures, like VGGNet, GoogLeNet, as well as ResNets, have recently demonstrated deeper CNN networks enabling greater performance but more discriminative features. The following are some of the important

TABLE 2 Layer configuration of HWNet v2.

CNN	L-1	L-2	L-3	L-4	ROI	FCNN
3 × 3	[3 × 3,64] × 3	[3 × 3,128] × 4	[3 × 3,256] × 6	[3 × 3], 3	[6 × 12]	[2,048] × 2

architectural innovations used in such networks that result in more efficient training: (a) using lower-dimensional filters (3 × 3), thus taking fewer parameters with large size filters that also perform as an essential regularizer, (b) by using (1 × 1) filters that perform as dimensionality lessening unit to retain the no. of parameters in rheostat, (c) by using inception layer that presents multi-scale functioning to have several equivalent layers functioning with different scales, and (d) using remaining blocks for studying remaining function $F(x) := H(x) - x$. Here $H(x)$ is anticipated for underlying mapping. The remaining layer is usually applied with a crosscut linking in the absence of available parameters.ResNet34 network of four blocks, each containing several ResNet modules in our upgraded HWNet architecture, dubbed HWNet v2, is used. Rather than employing globally averaged pooling, we discovered completely connected layers towards the conclusion of the process to learn superior features from the penultimate layer. The HWNet v2 network configuration is summarised in Table 2. Every ResNet module has two convolutional layers in it.

DenseNet models

The 121-layer DenseNet model is chosen as the basic model in this investigation since it produced the best results in most of the previous experiments on CXR classification. This model was initially trained by scratch, but we were able to improve its performance by using pre-learned weights using ImageNet. Binary Cross-Entropy (BCE), a multi-label loss, is used to train the model. We use the ImageNet dataset's mean as well as a standard deviation to normalise all the images. We apply data enhancement techniques like random horizontal flipping as well as uniformly random rotation from -15° to 15°.

To various image sizes, the Adam optimizer is applied at multiple primary acquiring rates. To varied image dimensions, the learning rate is measured by the actual batch size, which is shown in Table 3. If a validation loss does not reduce across three epochs, the learning rate decay is doubled to 0.5. If such a validation deficit cannot increase after ten epochs, the test must be terminated. Rather than using the MIMIC-CXR-JPG dataset's official train validation-test split, five-fold cross-validation is applied to 20% of test data from a dataset. The residual 80% of the data is being split as 90–10 for training validation purposes.

TABLE 3 Classification of learning rate.

Size of input image	Rate of learning	Group size of input image
256 × 256	0.0051	512
512 × 512	0.00034	256
1,024 × 1,024	0.000181	180
2,048 × 2,048	0.0000918	75

Effective receptive field

An effective receptive field (ERF) comprises the centre pixels inside the receptive field that have the maximum effect on the results. The impact of distribution inside the receptive field becomes asymptotically Uniform, according to Luo et al., and also, the effectual receptive field is just a segment of a theoretic receptive field. Moreover, as the network is taught, an effective receptive field grows larger. For calculating the ERF, a trained model is uploaded, run in evaluating manner, enters the data, and then averages the square gradients to get the uncentered sample variance. We use the centre pixel and backpropagation to generate one gradient per image. When computing the ERF, we also add huge random translations to the data. The value for image Xi at pixel (P, Q) is referred to as Xi (P, Q), as well as channel 'n', and pixel (x, y) of a NN's final convolutional layer activation is referred to as Fn (X_i, x, y). An ERF is a scalar image whose value shows the contribution of every pixel X_i (P, Q) to such activation. The partial derivative is used to calculate the impact, Eq. (8)

$$\frac{\alpha F^{a}(X_i,\ x,\ y)}{\alpha X_i(P,\ Q)} \tag{8}$$

This derivative is determined not just by the NN's weights but also by the input image X_i. Back-propagation is used for quickly calculating the partial derivative, Eq. (9)

$$ERF(P,\ Q) = \frac{1}{p,\ q}\sum_{p-1}^{p}\sum_{q-1}^{q}\frac{\alpha F^{a}(X_i,\ x,\ y)}{\alpha X_i(P,\ Q)} \tag{9}$$

where "n" denotes the number of links inside the network's final convolutional layer, and (x, y) denotes the central point for every image X_i. Since backpropagation gradients are positive and negative, first square them and then find the average for avoiding

cancellation, but the authors of the original ERF study do not square the gradients while averaging.

For various scale input images, we calculated the ERF from trained models. The ERF of DenseNet121 introduced models with different resolution photos is shown in Figure 8. With increased picture resolution, the ERF reduces in size in relation to the source image. When using CNNs upon high-resolution images, it suggests that large receptive fields are required.

Stacked scale-CNN model

We used a stacked CNN model in which discrete image scale sizes (256 × 256, 512 × 512, 1,024 × 1,024, and 2,048 × 2,048) are used to train the four models to enhance efficiency while keeping the flexibility for using pre-trained models. We train a specialised CNN model for deriving weights to each scale using the concatenation label probabilities of validation data from all these four classifiers providing features for a specific label. The CNN approach is depicted in Figure 9.

The weight to finding (F) and scaling (S) is wfs in our model of multi-label stacked CNN that possesses a single parameter P∈ $\mathcal{R}14 \times 4$, which is a 2-D tensor. In order to prove that a given label amount is one for the acquired weights of such four scales, we apply a SoftMax parametrization, Eq. (10)

$$w_{fs} = \frac{Exp(V_{fs})}{\sum_i Exp(V_{fs})} \tag{10}$$

For each finding f, its parameterization satisfies $\sum s\ w_{fs} = 1$. By defining $v_{fs} = 0$, we make the weights w_{fs} constant. These weights will then be used for averaging the probability of each scale.

Predicted label (PL):

$$\mathbf{PL_f\,(x_i) = \sum_i w_{fs} P_{fs}\,(x_i)} \tag{11}$$

We perform a numerically consistent calculation of $\sum s\ w_{fs}\sigma(\mathrm{PL}_{fs}(x_i))$ for each finding (F) for an image input of x_i with $\mathrm{PL}_{fs}(x_i)$, in which $\sigma(z) = (1 + \mathrm{Exp}(-R))1$ is the sigmoid function. Although all labels are calculated simultaneously in our approach, each label "F" is processed separately under this proposed model.

This stacking EM is trained using the validation data's output log produced by four classifiers with the help of four scales. The model is trained across 100 epochs with a 0.1 learning rate that decays exponentially, guaranteeing the converging of weights. For acquiring the ensemble probabilities for each label, multiply the learned weights by this model using predictions generated from test data.

Lion optimization algorithm

The lion optimization algorithm (LOA) is a bio-inspired algorithm that imitates lions' basic personality. Lions are naturally socialised, which causes them to divide into rover and habitant lions. A grown male lion is usually removed from the pride, and now these lions have no role in the pride, thus making them rovers. The rover lion can attack the lions inside the pride at some point in order to join the group. As a result, the LOA looks for rover lions and evaluates their quality. Fitness evaluation is the name given to this type of quality assessment, and only the fittest lions/lionesses are allowed to join the pride. The lions with lower fitness levels are omitted from the evaluation process. This cycle is repeated till better solutions are found. A certain percentage of lions are designated as rovers, while the others are assigned to points of pride. The gender ratio, as well as the border limit for each pride of lions, have been established.

The main key points of LOA are:

- In order to join the pride, the rover lion attempts to defeat the resident lion at any given time.
- The pride's fittest lion is the only one that can survive, and the rest unfit lions are ignored.

The overall number of points of pride remains constant, and rover lions have a chance of finding a proper fulfilment with a probability of [0,1].

Taking such benefits into consideration, this work uses LOA to determine the ideal filter order to denoising the image. When a filter is applied to a picture, the image's fitness is evaluated. The Peak Signal of Noise Ratio (PSNR) value is used to calculate the fitness value of such a work. The second filter is applied to the image once more, and the fitness value is checked. This method is repeated until the image does not quite improve for three consecutive rounds, at which point it is regarded as the best response. The following is the algorithm for the suggested task.

(A) Proposed Lion Optimization Algorithm
Step 1. The input of Noise-filled picture
Step 2. Denoised output Images
Step 3. Begin
Step 4. Establish a starting population of lions and prides
Step 5. Do
Step 6. Assess rover lions' matching;
Step 7. If matching (New)> matching (Old)
Step 8. Replace matching (Old);
Step 9. While (close condition);
Step 10. End

The PSNR value of an image is used to calculate image fitness, which analyses the quality between source

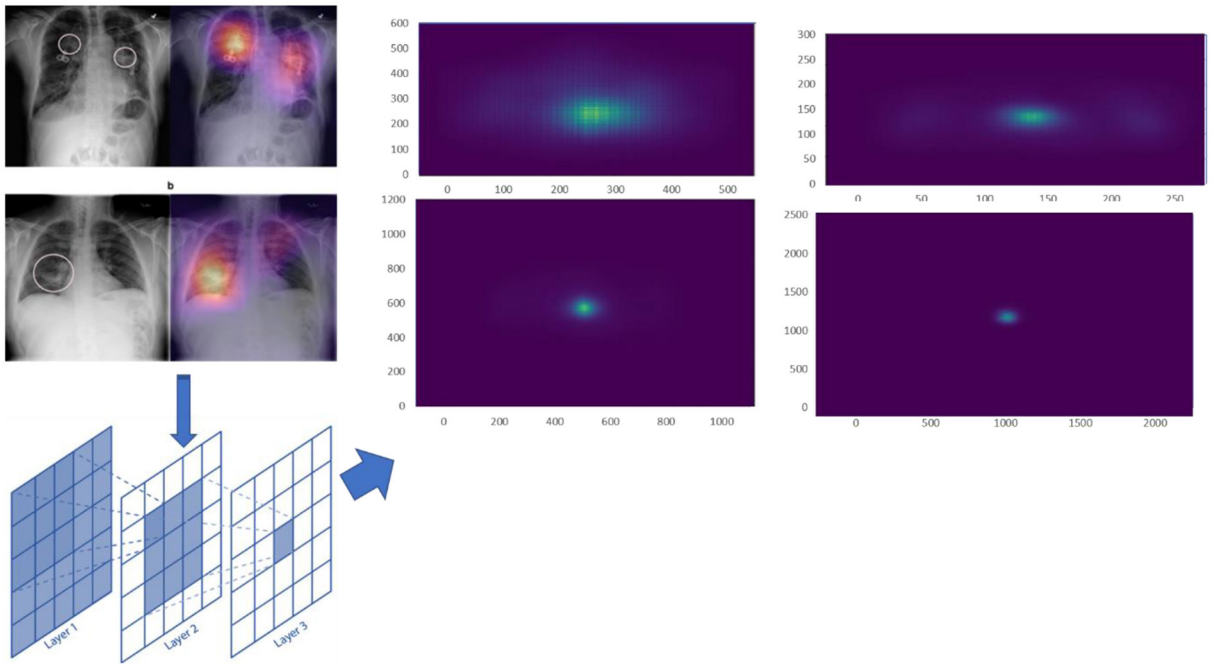


FIGURE 8
A trained Dense'et121's ERF.

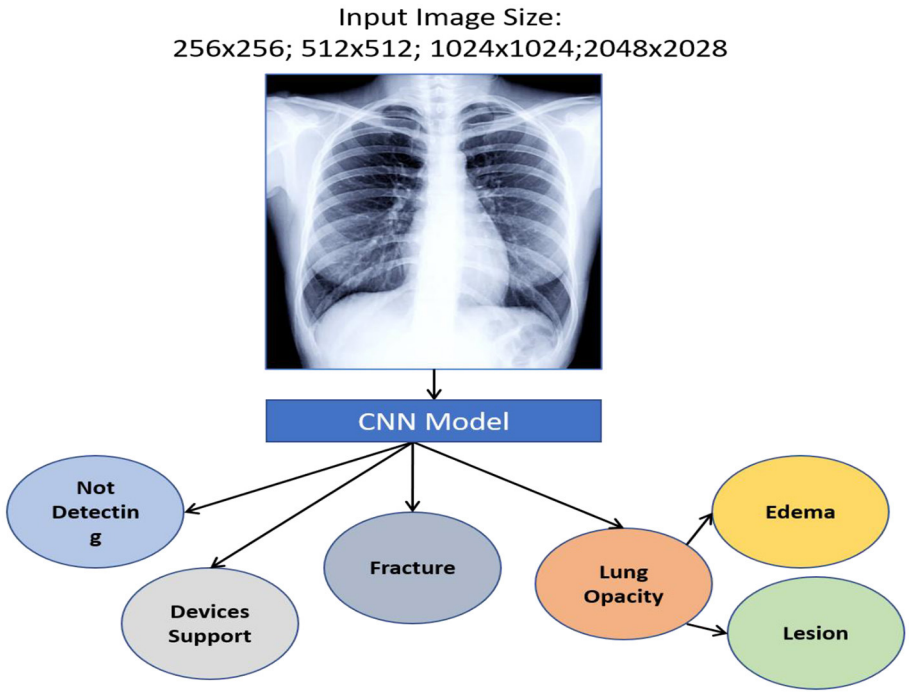


FIGURE 9
Workflow of CNN-based MedIMG.

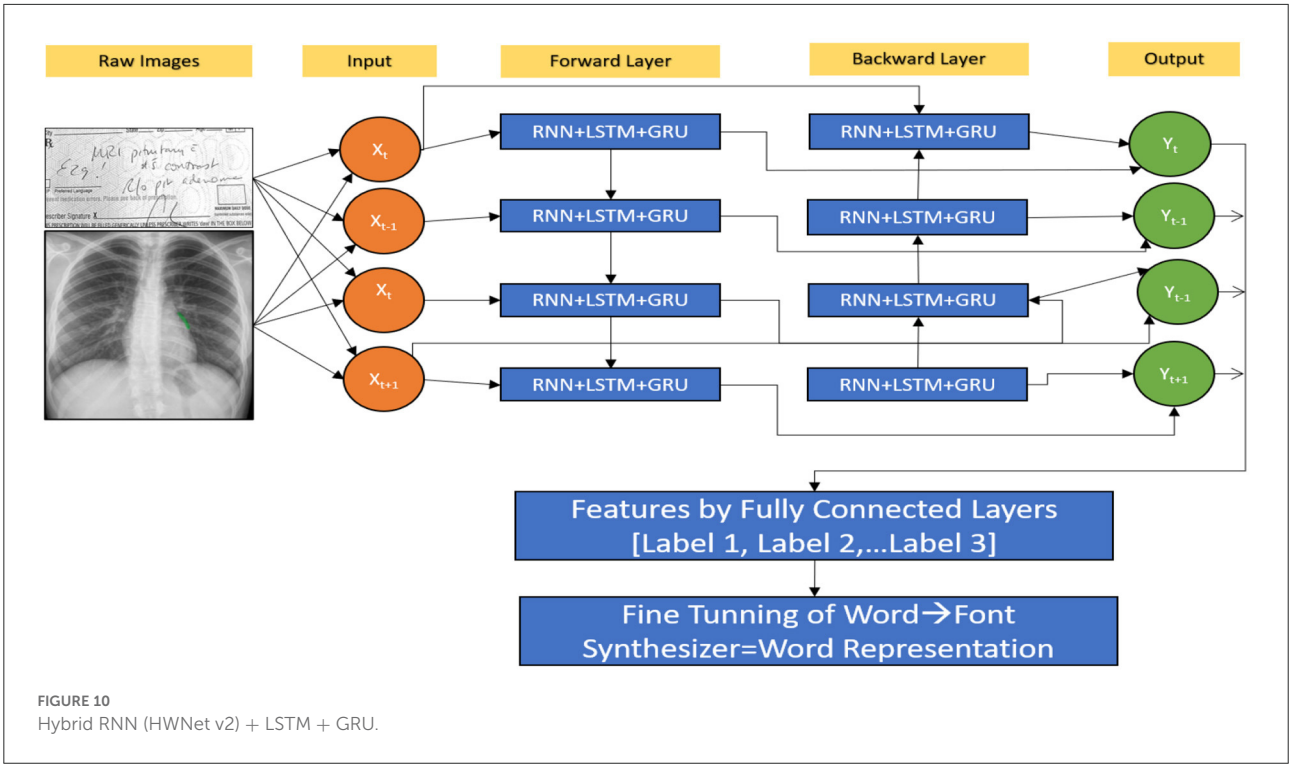


FIGURE 10
Hybrid RNN (HWNet v2) + LSTM + GRU.

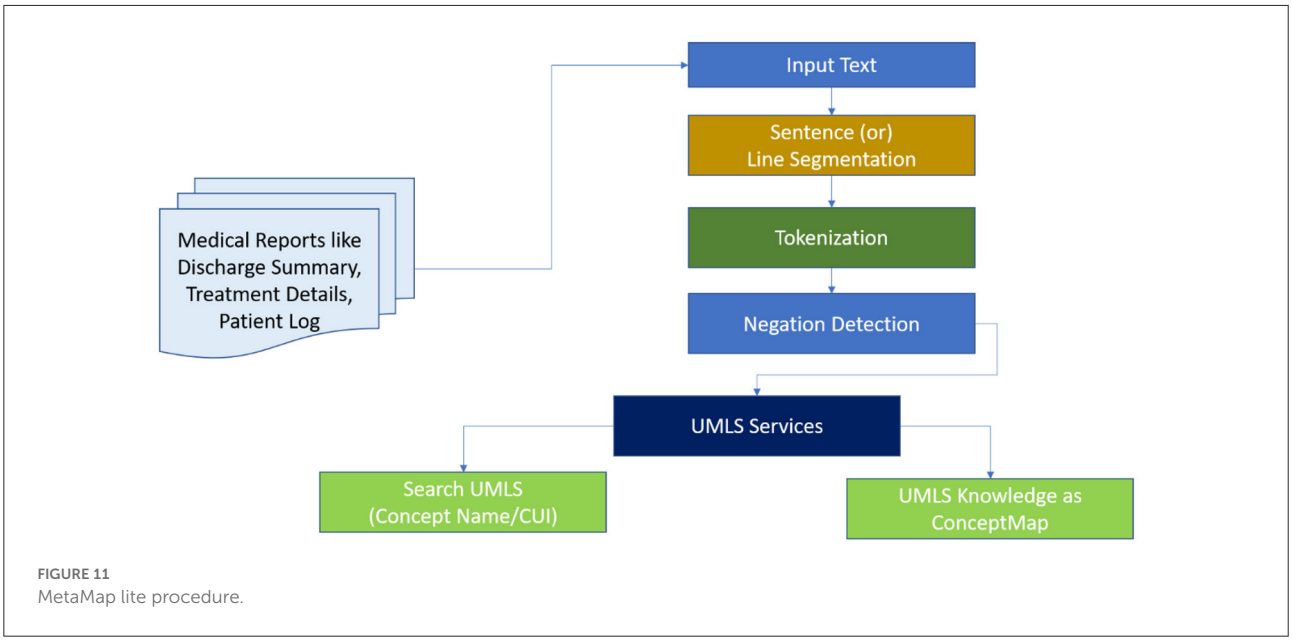


FIGURE 11
MetaMap lite procedure.

and denoised images. The following Eq. (11) is used to calculate this.

$$Suitable = 10 \times \log_{10} \frac{512 \times 512}{\frac{1}{hxw} \sum_{a-0}^{h-1} \sum_{b-0}^{w-1} [F(a,b) - G(a,b)]} DB \quad (12)$$

The height, as well as width of such images, are represented by h and w. The grey levels of the original MedIMG and denoised image at the (a, b) position are $F(a, b)$ and $G(a, b)$, respectively. The trial-and-error method is used to select all of the work's beginning settings. The PSNR value is used to calculate the

fitness, which allows the image quality to be evaluated and processed appropriately.

Proposed hybrid model to predictions of RNN (HWNet v2) + LSTM + GRM

The next character from MedIMG is predicted using an RNN. RNN's LSTM cell is used in this study. Each time LSTM processes a word, the following letter in the drug's name is determined by computing the probability of such possible values. The memory state of the network is initialised with a vector of '0's. So that it may be readily updated when each word has been scanned. The "x" in Figure 2 represents a feature, while the "o" represents a vector of size length probabilities. Because of their sequential structure, RNNs have gotten much attention for EHR predictive modelling. RNNs are Artificial Neural Networks that have recurrent connexions in the hidden layer. The HS of "ht" is updated successively based on activating the current input "xt" at a time "t," as well as the last HS of the layer ht−1. The Gated Recurrent Unit (GRU) and LSTM models are two popular RNN versions. The necessity for these variants arises because of the problems encountered while training traditional RNNs, where gradients would frequently vanish. Two major approaches have been investigated to address the vanishing gradient problem for extended sequences: (a) Develop an alternate training algorithm to the stochastic gradient, such as a clipped gradient. (b) Contain an internal cell state which controls the information flow *via* a series of gates in conventional RNN units. LSTMs and GRUs, unlike typical RNNs, have a recurrent internal loop at every unit and also with three and two gates to regulate information flow. Data have demonstrated that gated RNNs can identify long-term dependencies as well as avoid the problem of vanishing gradients.

A patient in an EHR is made up of a series of visits, and if each visit is represented as a numerical vector, those vectors can be fed into the RNN model in order to predict the desired outcome. In this study, we predominantly used the LSTM network to be the prediction model (Figure 10). The model recognises one visit for each time step of a patient "p" with "Tp" visits. From $t = 2$ to $t = \text{Tp}$, the new LSTM state is determined by the previous time step's state as well as the new input visit. The LSTM block spreads its state to subsequent dense layers at each time step. This indicates that the network has an input and output at each stage. Sequence-to-sequence prediction is the name given to this LSTM arrangement. We explored the multiple baseline prediction models, including several ML methods and LSTMs.

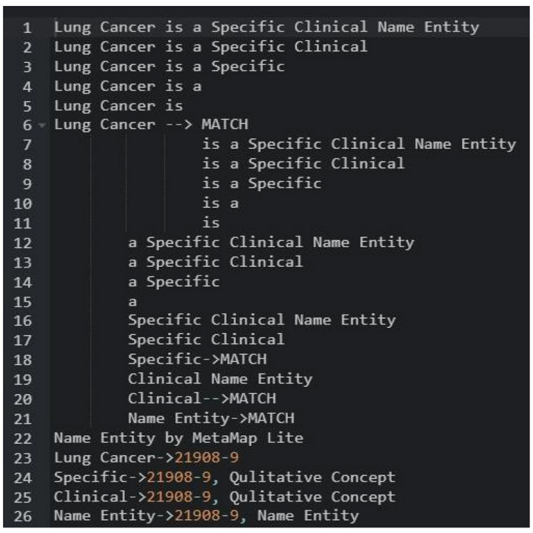


FIGURE 12
Test finding of UMLS-based NER.

NER-based open-source UMLS

MetaMapLite leverages ConText19/NegEx13 to negation detection. Consequently, the users who supply the longest concept appropriate to the semantic type of UMLS are provided. The processing pipeline is divided into seven steps, as shown in Figure 11 and detailed below.

Either the default OpenNLP sentence segmenter12 or its blank line segmenter is used for sentence/line segmentation. The basic MetaMapLite tokenization algorithm is used for tokenization. The part-of-speech tags that are optional are done by default OpenNLP part-of-speech tagger. The predefinition of the token window occurs for paralleling the sentence length using the OpenNLP segmenter; on using the blank line segmenter, a nonoverlapping window with 15 tokens occurs. There is no additional chunking. With minor modifications, term normalisation is based on MetaMapLite string normalisation. Before a dictionary lookup, the following procedures are performed on a term: (a) Parentheticals are removed, (b) syntactic inversion is performed, (c) lowercase is converted, and (d) possessives are stripped.

As shown in Figure 12, mapping is done to sentence/line-based chunks that are dynamically separated into sub-lists during processing. Each piece is normalised before being compared to a dictionary. Any dictionary match that is obliterated by a lengthier match gets discarded. MetaMapLite presently uses three dictionaries generated specifically for MetaMapLite for dictionary lookup: (1) CUI Concept, which maps CUIs to concept predefined names; (2) CUI Source Info, which holds the UMLS CUI, the UMLS string ID, a sequence ID, a source-derived string ID, a source abbreviation ID, and also a source term type ID; and (3) CUI St, that maps CUIs into semantic types. As an example of sentence-level NER, a token

TABLE 4 Test setup for hybrid model.

Image resolution	Execution time (s)	Memory (GB)	Image/s	GPU
256 × 256	30.10	10.39	250	4
512 × 512	112.90	21.39	75	3
1,024 × 1,024	267.18	35.56	25	5
2,048 × 2,048	1,024.48	38.38	7	6

TABLE 5 Assessment of word using mAP by hybrid model.

Training model (%)	1.0	0.9	0.8	0.7	0.6	0.3	0.0
Hybrid model	0.9567	0.9278	0.9187	0.93782	0.8919	0.88192	0.52762

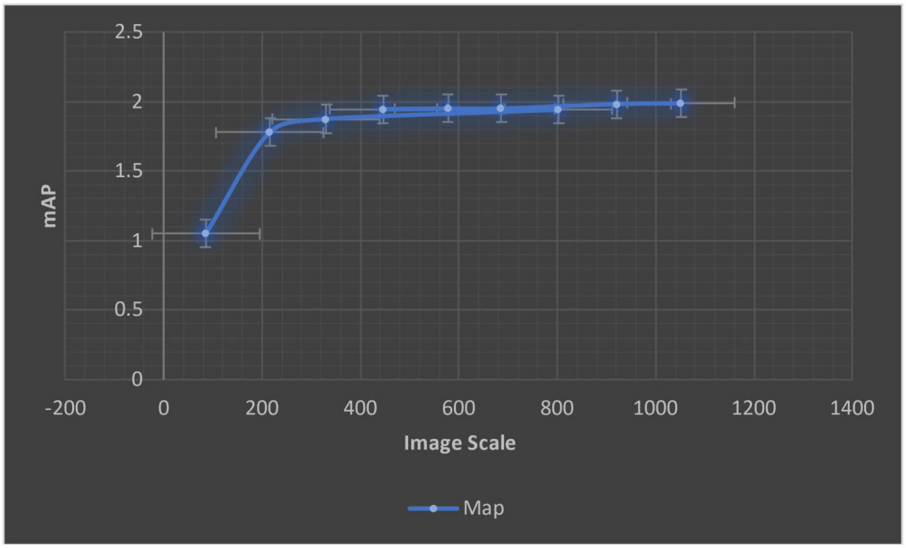


FIGURE 13
Performance analysis of learning hybrid model using MedIMG.

list for the phrase "Papillary thyroid cancer is a unique clinical object" is initiated and filtered.

Experimental results

The accuracy of a model can be improved by training it using a vast amount of data. As previously stated, we train the model using all generated data. TensorFlow's training is divided into two categories: iteration and model during training. We apply the TensorFlow standard model limiting iteration to around 10,000 for this research work. The dataset of MIMIC-CXR-JPG that was used in this investigation is discussed. It is a big one, comprising 377,110 X-rays and 227,827 MedIMG studies. It was created at the investigation level, which is an accumulation of photographs connected with a specific report.

As a result, many chest X-rays with frontal or lateral views can be included in a single scan. The dataset contains a sum of 2,50,100 frontal-view pictures and 1,20,300 lateral-view MedIMG. DICOM images are transformed into JPEGs in this collection. JPEG photos are not only less in size, but they are also easier to understand. The photos were de-identified before being converted to JPEG. After extracting picture pixels from the DICOM input file, the image pixels are normalised to the range [0, 255]. It was examined to determine if pixels are inverted and, if needed, images are inverted to ensure the image's greatest pixel appears as white and the low pixel as black. The contrast of the images was then improved using Contrast Limited Adaptive Histogram Equalisation (CLAHE). Finally, the photos are converted to JPEG format with a 95.78% quality factor and an 8-bit bit depth.

Computational setup

The NVIDIA DGX-A100 system, which consists of High Bandwidth Memory integrated with eight NVIDIA A100 40 GB Graphics Processing Units (GPUs), was utilised for training the DL models (HBM2). A total of 320 GB GPU memory and the highest level of system power consumption of 6.5kW are contained in this system. The code for our techniques was written using the Pytorch framework and the torch vision package. Table 4 shows a single GPU's consumption of memory for each resolution in gigabytes (GB) during training. As the solution improves, so does the amount of memory used. It depicts the computing capabilities and power required to carry out the experiments. Four A100 GPUs were used for a resolution of 256 × 256 and 512 × 512; however, six A100 GPUs are used simultaneously for 1,024 × 1,024 and 2,048 × 2,048 resolutions to speed up the training. As the image size is doubled, the number of trainable parameters increases, while the number of photos per second drops throughout validation.

A fascinating result of transfer learning using synthetic data. By changing the rate of training data associated with earlier studies, we test the reduced demand for real data for training the HWNet v2 model. We use the IAM dataset as our testbed again, but we are using a different composition of real data to compare performance with a model that uses all of the training data. Full is represented by 1.0, whereas 0.0 denotes the usage of just synthetic data. It is worth noting that all of these test results are trained on the completely synthetic data set before being fine-tuned with different proportions. As noticed, the performance reduction with less real training data begins slowly and only lowers by 6% when just 10% of genuine data is used. Although this Table 5 shows a lower reliance on real data, a more in-depth investigation (beyond the scope of this project) is required to assess the difference in domain space among synthetic data and handwritten target styles.

TABLE 6 Analysis of word on MedIMG.

Model	Supervision	Dataset 1	Dataset 2	Dataset 3
Hybrid	YES	0.9618	0.97182	0.9827

It is interesting to note that performance in the interval of 32–1,024 dimensions suffers little or no degradation, and we even notice a slight performance boost for 128 dimensions. This shows that the initial HWNet network may capture non-linear data relationships, but the final representing space exclusively contains linear components. On MedIMG with 32 sizes, we achieve an mAP of 0.8942, which is still state-of-the-art when compared to other related approaches shown in Table 6. It is worth noting that the performance attained with only eight dimensions (0.5942) remains superior to several non-DL techniques.

We want to compare the benefits of supervised vs. unsupervised learning, as well as compare performance on such datasets having in-depth features. Table 6 shows the results, which compare HWNet v2 (TPP) to previous techniques. For the English dataset, we apply the pre-trained model by IIIT-HWS with no fine-tuning; however, for the Hindi as well as Telugu datasets, we fine-tune their respective training corpus. The results show that word spotting performance has significantly increased on all of these datasets and that for printed English, HWNet v2 could be used off-the-shelf for various document tasks. The improved results in multi-languages show that architecture could be applied to a wide range of languages with different scripts and language elements. The qualitative outcomes displayed in the last two rows of Figure 13 also demonstrate the level of degrading in the retrieved words, indicating the efficiency of the suggested characteristics (Table 6).

Compression of embedding representation

There is inspiration from our prior studies on integrating the representation into a two-dimensional space, where we discovered surprising similarities between neighbours even under great compression. We now use Principal Component Analysis (PCA), a well-known linear dimensionality reduction technique, to explicitly obtain a lower-dimensional representation of HWNet characteristics. We retrieve the top eigenvectors within representation space from the validation data and use them to project each of the test data. On MedImg datasets, Figure 13 presents the performance difference across variable compression levels ranging from 4 to 1,024. Note that the original HWNet features are 2,048 bytes in size.

The outcome of hybrid model (RNN (HWNet v2) + LSTM + GRM)

The Hybrid model has the maximum area under the curve (AUC) for all 14 labels, as shown in Table 7. For each label, an AUC is at least 0.4% greater when compared with the outstanding performance of the single-scale model inclusive of the 1.1% average AUC, which is greater than the model of 1,024 × 1,024. This gap suggests that specialisation happens across scales to any given label, with certain examples best suited to other scales, so that the ensemble can be used to increase accuracy. Table 7 shows that the AUC values acquired for individual models closely resemble the learnt weights. The 1,024 × 1,024 scale, which seems to be the best performing unique model (bold), has been assigned the maximum weight

10.3389/fpubh.2022.926229

TABLE 7 AUCs for 15 TRAINED MedIMG.

ID	Finding	256 × 256	512 × 512	1,024 × 1,024	2,048 × 2,048	Hybrid model
1	Aches, pain, fever	81.8 ± 0.7	82.7 ± 0.2	81.9 ± 0.4	80.9 ± 0.5	82.4 ± 0.51
2	Addison disease	81.6 ± 0.4	81.67 ± 0.56	81.2 ± 0.5	79.9 ±0.4	82.6 ± 0.41
3	Adult brain tumours	82.1 ± 0.7	82.7 ± 0.2	82.5 ± 0.3	81.0 ± 0.2	83.5 ± 0.4
4	Agoraphobia	88.19 ± 0.3	89.18 ± 0.4	90.20 ± 0.2	89.23 ± 0.3	90.4 ± 0.3
5	Allergy	73.19 ± 0.9	73.8 ± 0.8	73.3 ± 0.8	72.3 ±1.1	74.8 ± 1.5
6	Anaemia	67.7 ± 1.9	67.14 ± 1.4	67.23 ± 1.3	65.6 ± 0.9	68.7 ± 1.7
7	Dandruff vs. dry scalp	73.8 ± 0.9	75.0 ± 0.9	74.5 ± 0.7	73.8 ± 0.5	77.0 ± 0.9
8	Diabetes insipidus	74.9 ± 0.4	76.1 ± 0.2	76.3 ± 0.3	75.4 ± 0.4	77.0 ± 0.4
9	Dialysis	85.3 ± 0.4	85.8 ± 0.4	85.9 ± 0.3	85.3 ± 0.2	86.5 ± 0.15
10	Eczema facts	91.9 ± 0.4	92.3 ± 0.4	92.3 ± 0.3	91.5 ± 0.4	92.8 ± 0.5
11	Calcific bursitis	80.6 ± 0.3	82.5 ± 1.1	82.2 ± 0.7	81.6 ± 0.5	84.0 ± 0.8
12	Cardiac catheterization	71.4 ± 0.6	72.9 ± 0.4	73.1 ± 0.5	71.6 ± 0.6	74.3 ± 0.5
13	Capsule endoscopy	85.6 ±1.1	87.91 ± 0.718	88.4 ± 0.9	88.7 ± 0.4	90.0 ± 0.7
14	Chest X-ray	89.8 ± 0.4	91.8 ± 0.1	92.13 ± 0.3	92.1 ± 0.2	92.7 ± 0.3
15	Lung cancer	80.5 ± 0.3	81.5 ± 0.3	81.25 ± 0.3	80.17 ± 0.12	82.6 ±0.3

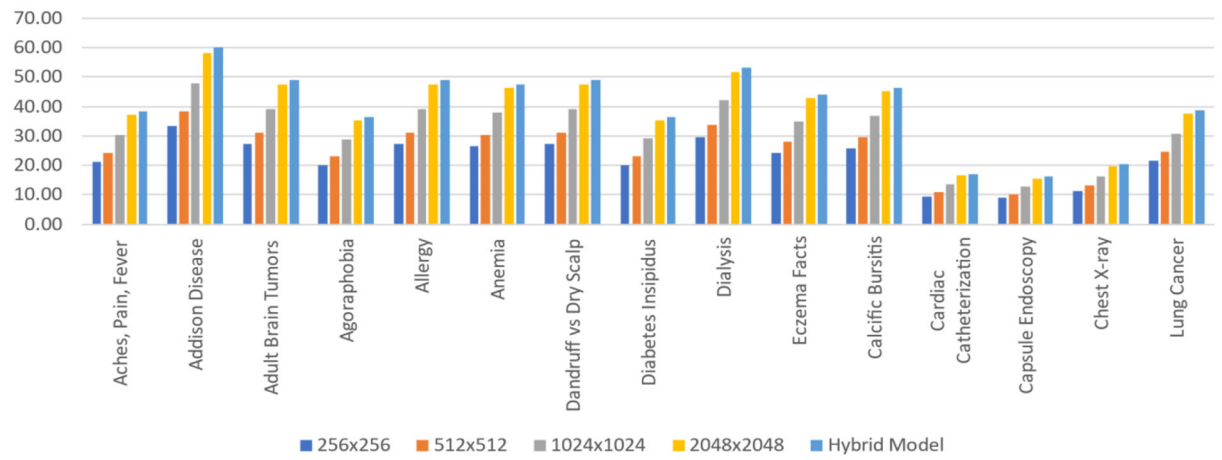


FIGURE 14
Average stacked scale CNN hybrid model.

(32.7%) in computing prediction for "Aches, Pain, Fever" with the Hybrid model. Ten of 15 tasks fall under this category.

This study evaluates model performance using the AUC metric. The area beneath the Receiver Operating Characteristic (ROC) curve (38) is measured by AUC (39). True Positives (TP) and True Negatives (TN) are improved by a model with a higher AUC (40). Table 3 shows the AUC scores for all 15 labels, including the average, for various scales. Because, throughout training, the application of five-fold cross-validation was prevalent. Those AUC scores are the AUC scores average from each of the five folds. Throughout the standard deviations calculated over the five-folds, the AUC scores are displayed. As per our hypothesis, the AUC score for all labels should grow as the input image size increases, i.e., all the bold values should be in 2,048 × 2,048. However, as the table shows, it is not a fact. Some labels meet up to our expectations, while others do not. The AUC score for "Cardiac Catheterization," for example, drops as image size increases. When compared to 256 × 256 resolution, a few labels performed well with 512x512 resolution, but their AUC scores drop when the resolution is raised to 1,024 × 1,024 resolution.

For each of the 15 tasks, Figure 14 represents the layered Hybrid Model's learnt weight distribution on four scales. When

calculating the predictions for the Hybrid Model, this bar chart displays that resolution is assigned more weights to a specific task. It is worth noting that such weights are given in percentages, and for each task, the weights from all five scales add up to 100%.

Implementation of MetaMap lite

Table 8 shows the F1scores, Recall, and Precision [42, 43] on every tool to every collection that we are capable of proceeding with it. Because negative detection isn't a natural MetaMapLite feature, and just comparing this one with MetaMapLite implementations of NegEx is to make sure of not losing any performance. We were unable to assess DNorm on two datasets due to difficulty in getting relevant offsets.

TABLE 8 Test report of feature extraction using MetaMapLite.

MetaMapLite	Model
3 m 31.89 s	Aches, pain, fever
2 m 21.17 s	Addison disease
1 m 34.17 s	Adult brain tumours
2 m 24.28 s	Agoraphobia
3 m 25.38 s	Allergy
2 m 17.38 s	Anaemia
3 m 34.48 s	Dandruff vs. dry scalp
2 m 13.41 s	Diabetes insipidus
4 m 21.38 s	Dialysis
3 m 24.78 s	Eczema facts
2 m 17.89 s	Calcific bursitis
4 m 23.67 s	Cardiac catheterization
2 m 19.89 s	Capsule endoscopy
4 m 18.79 s	Chest X-ray
3 m 23.78 s	Lung cancer

We made considerable effort to install, run, and collect mention-level offsets to third-party programmes. However, with the versions of DNorm available upon this i2b2 2010 and ShARe collections, it cannot obtain mention-level data. It is reasonably sure that the findings we obtained are comparable with those obtained by average end-users of versions of such tools present at the time we conducted our research investigations. On all collections, MetaMapLite received F1scores that were greater than any other tool. The absolute results achieved in our trials may vary from those reported previously, which could be due to differences in tool versions and settings used in our experiments. On request, we may provide all of our tool installations and test collections, which can be made accessible for evaluation scripts. MetaMapLite also takes various restrictions (Figure 15).

It only implements a small portion of the rich MetaMapLite options. MetaMapLite is analysed with the default MetaMap setting, and its performance is equivalent to out-of-the-boxMetaMapLite setups that haven't been tuned for a particular activity.

The processing speed of MetaMapLite is determined by the options selected: the quickest processing is accomplished without POS tagging by a negligible drop in the F-score.

MetaMapLite does not currently have any word-sense disambiguation modules, although it will offer all UMLS senses for a given term. Roughly, mappings that are usually incorrect might be put to a rejection output, similar to MetaMapLite; otherwise, a custom-made data file is handled as a substitute for the tools' regular lexicon records. Finally, the examination has a significant restriction in that it only examines disorders and denials.

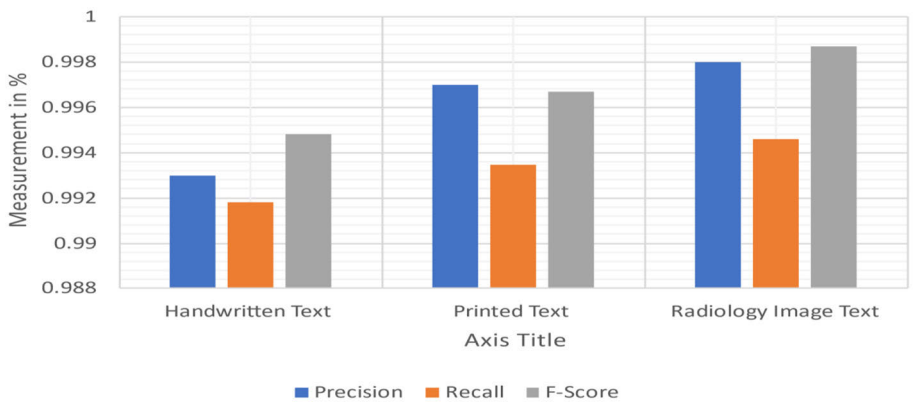


FIGURE 15
Measuring recall, precision, and F1-scores.

Conclusion

In this paper, Deep Learning is used to extract knowledge from a clever approach to distinguish the doctor's prescription photos and radiology images. The main outcomes are proof of ML methods' ability to discern which elements in short free-text clinical prescriptions are considered inclusion/exclusion criteria, as well as a huge radiology image dataset of images, metadata, radiological results, diagnoses, and anatomic locations. We provide a generic DCNN system for learning document image word image representation. The Hybrid DL Model's core architecture utilises synthetic data to rapidly pre-train multiple data expansion strategies to replicate a typical mechanism at document text formation. Some new visualisation approaches are used to provide other views into the fine-tuning process and to comprehend the differences learned at various levels. On challenging historical texts in both handwritten and radiology modalities, we successfully demonstrate the stability of learned representation in aspects of both performances as well as the dimensionality of the final word. The significance of feature and image resolution in CNN design enabling automated MedIMG categorisation is demonstrated in this study. MetaMapLite, a lightweight software application using MetaMap, is among very widely applied termed NER technique to identify Metathesaurus UMLS ideas as texts on bio-medicine, is described herein our research investigation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualisation and funding acquisition: SB and MA. Investigation and methodology: PD and SB. Project administration: SB. Resources: MA. Writing of the original draft: SS and PD. Writing of the review and editing: SB and SS. Validation and visualisation: PD. Data curation: MA and PD. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. 159].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Bhattacharya S, Maddikunta PK, Hakak S, Khan WZ, Bashir AK, Jolfaei A, et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimed Tools Appl.* (2020) 9:1–25. doi: 10.1007/s11042-020-09988-y

2. Iwendi C, Moqurrab SA, Anjum A, Khan S, Mohan S, Srivastava G. N-sanitization: a semantic privacy-preserving framework for unstructured medical datasets. *Comput Commun.* (2020) 161:160–71. doi: 10.1016/j.comcom.2020.07.032

3. Iwendi C, Ponnan S, Munirathinam R, Srinivasan K, Chang CY. An efficient and unique TF/IDF algorithmic model-based data analysis for handling applications with big data streaming. *Electronics.* (2019) 8:1331. doi: 10.3390/electronics8111331

4. Panayides AS, Amini A, Filipovic ND, Sharma A, Tsaftaris SA, Young A, et al. AI in medical imaging informatics: current challenges and future directions. *IEEE J Biomed Health Inform.* (2020) 24:1837–57. doi: 10.1109/JBHI.2020.2991043

5. Li Z, Hubchak M, Zhu Y. Deep leakage from gradients in multiple-label medical image classification. In: *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI).* Victoria, BC: IEEE (2021). p. 447–8. doi: 10.1109/ICHI52183.2021.00078

6. Noothout JM, De Vos BD, Wolterink JM, Postma EM, Smeets PA, Takx RA, et al. Deep learning-based regression and classification for automatic landmark localization in medical images. *IEEE Trans Med Imaging.* (2020) 39:4011–22. doi: 10.1109/TMI.2020.3009002

7. Abbas S, Alhwaiti Y, Fatima A, Khan MA, Khan MA, Ghazal TM, et al. Convolutional neural network based intelligent handwritten document recognition. *Comput Mater Contin.* (2022) 70:4563–81. doi: 10.32604/cmc.2022.021102

8. Ahn E, Kumar A, Feng D, Fulham M, Kim J. Unsupervised deep transfer feature learning for medical image classification. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019).* Venice: IEEE (2019). p. 1915–8. doi: 10.1109/ISBI.2019.8759275

9. Pang S, Pang C, Zhao L, Chen Y, Su Z, Zhou Y, et al. SpineParseNet: spine parsing for volumetric MR image by a two-stage segmentation framework with semantic image representation. *IEEE Trans Med Imaging.* (2021) 40:262–73. doi: 10.1109/TMI.2020.3025087

10. Liang C, Xin S. Research status and prospects of deep learning in medical images. In: *2020 International Conference on Communications, Information System and Computer Engineering (CISCE).* (2020), p. 380–2. doi: 10.1109/CISCE50729.2020.00084

11. Xue D, Zhou X, Li C, Yao Y, Rahaman MM, Zhang J, et al. An application of transfer learning and ensemble learning techniques for

cervical histopathology image classification. *IEEE Access.* (2020) 8:104603–18. doi: 10.1109/ACCESS.2020.2999816

12. Shayesteh SP, Nazari M, Salahshour A, Avval AH, Hajianfar G, Araabi M, et al. Machine learning-based malignancy prediction in thyroid nodules malignancy: radiomics analysis of ultrasound images. In: *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. Boston, MA: IEEE (2020), p. 1–2. doi: 10.1109/NSS/MIC42677.2020.9507959

13. Yu Y, Favour E, Mazumder P. Convolutional neural network design for breast cancer medical image classification. In: *2020 IEEE 20th International Conference on Communication Technology (ICCT)*. (2020), p. 1325–32. doi: 10.1109/ICCT50939.2020.9295909

14. Ma Y, Liu J, Liu Y, Fu H, Hu Y, Cheng J, et al. Structure and illumination constrained GAN for medical image enhancement. *IEEE Trans Med Imaging.* (2021) 40:3955–67. doi: 10.1109/TMI.2021.3101937

15. MacGillivray TJ, Cameron JR, Zhang Q, El-Medany A, Mulholland C, Sheng Z, et al. Suitability of UK Biobank retinal images for automatic analysis of morphometric properties of the vasculature. *PLoS ONE.* (2015) 10:1–10. doi: 10.1371/journal.pone.0127914

16. Liu C, Hsu M, Wang L, Wen L, Lai J. A unified approach to adoption of laboratory LOIN in Taiwan. In: *2007 9th International Conference on e-Health Networking, Application and Services*. Taipei (2007). p.144–9. doi: 10.1109/HEALTH.2007.381620

17. Eilbeck K, Jacobs J, Staes CJ. Optimize querying of LOINC® with an ontology: give me the chlamydia tests the epidemiologists want me to use! In: *2013 46th Hawaii International Conference on System Sciences*. (2013), pp. 2398–407. doi: 10.1109/HICSS.2013.433

18. Dalmiani S, Morales MA, Mazzarisi A, Marcheschi P, Carpeggiani C. EMR standards in cardiological outpatient management. *Comput Cardiol.* (2004) 2004:613–6. doi: 10.1109/CIC.2004.1443013

19. Iwendi C, Khan S, Anajemba JH, Mittal M, Alenezi M, Alazab M. The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems. *Sensors.* (2020) 20:2559. doi: 10.3390/s20092559

20. Kaiser K, Gschwandtner T, Martini P. MapFace - an editor for MetaMap transfer (MMTx). In: *2008 21st IEEE International Symposium on Computer-Based Medical Systems*. Jyvaskyla: IEEE (2008). p. 150–2. doi: 10.1109/CBMS.2008.116

21. Pyae Khin NP, Lynn KT. Medical concept extraction: a comparison of statistical and semantic methods. In: *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. Kanazawa: IEEE (2017). p. 35—8. doi: 10.1109/SNPD.2017.8022697

22. Yu Y, Li M, Liu L, Wu F-X, Wang J. Tentative diagnosis prediction *via* deep understanding of patient narratives. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (2019), p. 1000–3. doi: 10.1109/BIBM47256.2019.8983129

23. Zexian Z, Ankita R, Xiaoyu L, Sasa E, Susan C, Seema K, et al. Using clinical narratives and structured data to identify distant recurrences in breast cancer. In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. (2018), p. 44–52. doi: 10.1109/ICHI.2018.00013

24. Luque C, Luna JM, Ventura S, MiNerDoc: a semantically enriched text mining system to transform clinical text into knowledge. In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. (2019), p. 702–7. doi: 10.1109/CBMS.2019.00142

25. Rui X, Cao X, Xie Q, Yue Z, Zhao Q, Meng D. Learning an explicit weighting scheme for adapting complex HSI noise. In: *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. (2021), p. 6735–44. doi: 10.1109/CVPR46437.2021.00667

26. Marinai S, Gori M, Soda G. Artificial neural networks for document analysis and recognition. *IEEE Trans Pattern Anal Mach Intell.* (2005) 27:23–35. doi: 10.1109/TPAMI.2005.4

27. Krishnan P, Jawahar CV. 2019 HWNet v2: an efficient word image representation for handwritten documents. *Int J Doc Anal Recognit*. (2019) 22:387–405. doi: 10.1007/s10032-019-00336-x

28. Stromer D, Christlein V, Maier A, Zippert P, Helmecke E, Hausotte T, et al. Non-destructive digitization of soiled historical Chinese bamboo scrolls. In: *Proceedings of the 13th IAPR International Workshop on Document Analysis Systems*. Vienna: DAS (2018), p. 55–60. doi: 10.1109/DAS.2018.37

29. Yin X, Aldarrab N, Megyesi B, Knight K. Decipherment of Historical Manuscript Images. In: *Proceedings of the 2019 International Conference on Document Analysis and Recognition*. Sydney: ICDAR. (2019), p. 78–85. doi: 10.1109/ICDAR.2019.00022

30. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Salt Lake City, UT (2018), p. 4510–20. doi: 10.1109/CVPR.2018.00474

31. Tensmeyer C, Morariu VI, Price BL, Cohen S, Martinez TR. Deep splitting and merging for table structure decomposition. In: *Proceedings of the 2019 International Conference on Document Analysis and Recognition*. Sydney: ICDAR (2019), p. 114–21 doi: 10.1109/ICDAR.2019.00027

32. Yepes J. Aronson AR. Query expansion for UMLS metathesaurus disambiguation based on automatic corpus extraction. In: *2010 Ninth International Conference on Machine Learning and Applications.* Washington, DC (2010). p. 965–8. doi: 10.1109/ICMLA.2010.154

33. Uddin I, Ramli DA, Khan A, Bangash JI, Fayyaz N, Khan A, Kundi M. Benchmark Pashto handwritten character dataset and Pashto object character recognition (OCR) using deep neural network with rule activation function. *Complexity*. (2021) 2021:6669672; 16. doi: 10.1155/2021/6669672

34. Shah V, Shah B, Goswami R, Kumar S, Moradiya C. Creation of unambiguous centralized knowledge base from UMLS metathesaurus. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. (2018), p. 1270–6. doi: 10.1109/BIBM.2018.8621384

35. Bhalla K, Koundal D, Bhatia S, Khalid M, Rahmani I, Tahir M. Fusion of infrared and visible images using fuzzy based Siamese convolutional network. *Comput Mater Con*. (2022) 70:5503–18. doi: 10.32604/cmc.2022.021125

36. Schäfer H, Friedrich CM. UMLS mapping and Word embeddings for ICD code assignment using the MIMIC-III intensive care database. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin (2019). p. 6089–92. doi: 10.1109/EMBC.2019.8856442

37. Li R, Wang X, Liu Y, Zhang S. Improved technology similarity measurement in the medical field based on subject-action-object semantic structure: a case study of Alzheimer's disease. In: *IEEE Transactions on Engineering Management*. doi: 10.1109/TEM.2020.3047370

38. Lakshmanna K, Kaluri R, Reddy GT, Nagaraja G, Subramanian DV. An enhanced algorithm for frequent pattern mining from biological sequences. *Int J Pharm Technol.* (2016) 8:12776–84.

39. Dhanamjayulu C, Nizhal UN, Maddikunta PKR, Gadekallu TR, Iwendi C, Wei C, et al. Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning. In: *IET Image Processing*. (2021).

40. Bhatia S, Alam S, Shuaib M, Alhameed MH, Jeribi F, Alsuwailem RI. Retinal vessel extraction *via* assisted multi-channel feature map and U-Net. *Front Public Health*. (2022) 10:858327. doi: 10.3389/fpubh.2022.858327

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fpubh.2022.944605

Check for updates

OPEN ACCESS

EDITED BY
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

REVIEWED BY
Qiang Wang,
Wuhan University, China
Luobing Dong,
Xidian University, China

*CORRESPONDENCE
Xiaoting Wang
sujiang202205@163.com

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 15 May 2022
ACCEPTED 27 June 2022
PUBLISHED 09 September 2022

CITATION
Liu F and Wang X (2022) Influence of seamless nursing model of humanistic care on nursing quality and blood infection risk of neurosurgical ICU patients and its correlation with APACHE II score.
Front. Public Health 10:944605.
doi: 10.3389/fpubh.2022.944605

COPYRIGHT
© 2022 Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Influence of seamless nursing model of humanistic care on nursing quality and blood infection risk of neurosurgical ICU patients and its correlation with APACHE II score

Fuhua Liu[1] and Xiaoting Wang[2]*

[1]Neurosurgery Department, Zhejiang Hospital, Hangzhou, China, [2]Intensive Care Unit Department, Zhejiang Hospital, Hangzhou, China

This research aims to explore the influence of seamless nursing mode of humanistic care on nursing quality and blood infection risk of ICU patients in neurosurgery, and the model of correlation with APACHE II score. 110 ICU patients are grouped into control set and study set, which are, respectively, given the previous routine nursing and the seamless management based on humanistic care to compare the two sets in the following aspects: nursing quality, blood infection rate, anxiety and depression extension before and after nursing, nursing satisfaction and APACHE II score, and to figure out the correlation between patient nursing quality score, and to compare blood infection and APACHE II score. Comparison and statistical analysis are used to disclose the influence and the correlation. The results show that there is not only a large negative correlation between nursing quality scores and APACHE II scores, but also a large negative correlation between the risk of blood infection and APACHE II score.

KEYWORDS

humanistic care, seamless nursing, neurosurgical ICU, nursing quality, blood infection, APACHE II score

Introduction

In the ICU treatment of neurosurgery, most patients are often confronted with the characteristics of disease severity and rapid variance in the disease so it is necessary to provide patients with timely and effective care to ensure their health and life. However, the number of ICU patients in neurosurgery and the heavy workload force nurses to undertake high psychological pressure, leading to a series of problems and errors in nursing, which, consequently, have side effects on the efficacy and prognosis (1). For neurosurgical ICU patients, most of them have gone through surgical treatment and may stay in the situation of severe physiological dysfunction and organ failure, not out of the danger period so they need further monitoring and treatment. However, surgery invasive operations such as trauma and tracheal intubation, which is required by the

diseases themselves, increase the infection risk to some extent and have a serious impact on the prognosis and recovery. It is important to make the prevention and control of nosocomial infection in the neurosurgical ICU. And there is a need to find a safe and reliable method. It is one of the key clinical processes at present (2, 3). With the medicine development and the advancement in medical technology, the nursing model has been continuously updated and developed. Humanistic care is a human-oriented nursing model, and its application in the process of ICU nursing management can help patients increase compliance (4) and establish confidence in overcoming the disease. The seamless management is one of new models in modern nursing. The concept is based on modernization, which can largely ensure the integrity and effectiveness of nursing services (5, 6). However, the application of humanistic care combined with seamless management in neurosurgical ICU patients is rarely seen in present clinical practice, and there is no strong and effective basis for evaluating its efficacy. The impact of the seamless care model on the care quality and blood infection risk in neurosurgical ICU patients will now be explained in the following. Humanistic care in the treatment of critically ill patients can make nursing staff fully respect the patient's personality and human rights and meet the reasonable needs of patients. This nursing model can not only provide the humanistic care required for overall nursing, but also support the improvement of the professional quality of nursing staff. Therefore, humanistic care is the new era for nursing staff.

In this paper, 110 neurosurgical ICU inpatients admitted to our hospital from September 2020 to January 2021 are selected for research, and they are grouped into control set and study set according to the sequence of admission, with 55 cases in each set. Among them, there are 26 patients diagnosed with underlying diseases and 29 patients without underlying diseases; there were 30 males and 25 females in the research set, with an age range of 39∼68, average age being 56.62 ± 4.75 for the male patients. In the routine set, there were 32 male cases, and 23 cases were female, the age range being 40–69 years old. The average age was 57.19 ± 4.86. There were 28 cases with underlying diseases, 27 cases without underlying diseases. For general clinical data such as disease, $P > 0.05$, there was no large difference between the sets. All patients signed the informed consent form, and this study was approved by the Medical Ethics Committee in our hospital.

This paper finds that the nursing quality scores of the study set, including theory, practice and comprehensiveness, were notoriously higher than those of the control set; the risk rate of blood infection during hospitalization in the study set was 5.45%, which was notoriously lower than that of the control set, 25.45%. After nursing, the SAS and SDS scores of the two sets decreased, and the research set was lower than the control set; after nursing, the APACHE II scores of the two sets were improved, and the improvement effect of the research set was notoriously better than that of the control set ($P < 0.05$); the seamless nursing model based on humanistic care can improve nurses' service attitude, improve doctor-patient relationship, provide patients with appropriate psychological care, correct patients' misunderstandings, create a positive attitude, eliminate negative emotions such as fear and anxiety, and make them treat the disease with optimism and increase the confidence in treatment. Additionally, this study also compared the nursing satisfaction of the two sets. The results found that the total nursing satisfaction of the research set was 89.09%, which was notoriously higher than that of the control set (63.64%) ($P < 0.05$). For neurological ICU patients, seamless nursing mode based on humanistic care can help improve the quality of care, reduce the risk of blood infection and APACHE II score, promote the recovery of patients, improve the prognosis of patients, and present clinical application potential.

The rest of this paper is organized as the follows: Section *Related work* discusses the related work, followed by the proposed methods in Section *Our proposed method*. Section *Experimental results and analysis* is the content of results in which the contrast between the main factors is conducted, and Section *Conclusion and the future work* concludes the paper with summary and future research directions.

Related work

In recent years, with the continuous improvement of people's living standards, people's demand for medical care is increasing too, especially for the quality of patient care in ICU, which is directly related to the health of patients. The intensive care unit is an important place to provide emergency care for critically ill patients so a comprehensive, systematic, scientifically and professional nursing is required to provide a safer guarantee for the lives of patients (7). Therefore, strengthening effective nursing measures in the treatment of ICU patients is very important for the treatment and recovery of ICU patients. Seamless nursing management is a method that mainly includes the detailed understanding and management of the patient's condition, and the development of a targeted nursing plan (8). The good nursing management system ensures the smooth nursing process, thereby effectively improving the quality of nursing, reducing the incidence of patient infection, and enabling ICU patients to receive high-quality and safe nursing services during the treatment process. Humanistic care nursing is based on the principle of "human-oriented", which can show more emotional respect and love for patients while carrying out routine nursing (9). The main reasons for the analysis are that after nutritional support based on the seamless nursing model based on humanistic care, the patient's immunity can be notoriously improved; the standardized application of antibiotics can effectively prevent superinfection and the formation of drug-resistant bacteria; strengthening environmental management and oral care can prevent respiratory infections; good disinfection of the urethra and catheter can effectively reduce the risk of infection and

avoid urinary tract infections by closely monitoring the situation of the surgical incision and puncture site, and changing the bandage in time, as well as helping the patients turn over regularly, to effectively prevent skin and soft tissue infections (10, 11). Romero-Garcia et al. (12) explored the application value of seamless management model for nosocomial infection in surgical ICU, and found that providing appropriate seamless management nursing intervention can notoriously reduce the possibility of nosocomial infection in patients, which is consistent with some results of this study. For most ICU doctors and nurses, most of the time is highly stressful. In this regard, hospitals should actively take appropriate measures to reduce the pressure on nursing staff and provide better services for patients (13). The seamless nursing model focuses on scientific management, creating a seamless nursing management team so that ICU patients can create a rigorous, continuous, and integrated seamless nursing process before admission, during treatment, and after treatment (14, 15).

The application of seamless nursing management strengthens nursing management, makes nursing measures practical, improves the awareness and motivation of nursing staff, and enhances the sense of responsibility of nursing staff. At the same time, nursing responsibility can notoriously improve the quality of nursing through perfect supervision (16). It is a new nursing model that is welcomed by patients and recognized by the society (17). The application of seamless management can effectively care for the psychology of patients and improve the bad mood of patients, enable nursing staff to communicate well with patients, improve patients' confidence in nursing staff, and reduce the occurrence of medical disputes, with employees' awareness of risk prevention greatly enhanced, preventing risk events before they occur, controlling every reference and detail in nursing work, and notoriously reducing the incidence of nursing risk events while providing patients with safe, high-quality, and comprehensive nursing services (18, 19).

Our proposed method

For the selection of patients, there are two standards including inclusion and exclusion.

The inclusion criteria include the following aspects: (1) All patients were neurosurgical ICU patients; (2) The selected patients were over 18 years old; (3) The patients had barrier-free consciousness and could communicate and communicate with doctors normally; (4) All patients signed informed consent notice.

The exclusion criteria include the following aspects: (1) Patients with poor treatment compliance and unable to determine the curative effect; (2) Patients with severe illness requiring emergency treatment; (3) Patients with severe complications; (4) Patients who have used contraindicated drugs; (5) Patients with missing data on cases and related impact examinations; (6) Patients' condition deteriorated suddenly during the research process; (7) Patients with other diseases affected the research results; (8) Patients who had unexpected circumstances during the research process, made it difficult to continue the research; (9) Serious disease during the research process affected other researchers.

Control set

The set of patients can use the previous routine nursing mode, and the details are as follows: (1) closely monitor and observe the patient's condition and vital signs, evaluate the patient's physical responses according to the patient's actual situation, and give appropriate medication or (2) during the nursing process, closely observe and record the changes in the patient's condition, and carry out routine management and monitoring; (3) if the patient's condition improves, appropriate measures should be taken, and instruct patients to carry out relevant rehabilitation exercises, and communicate with the patient's families to guide some precautions after surgery, and provide necessary health guidance or advice to reduce the incidence of postoperative complications.

Study set

Based on the nursing in the control set, the patients in the set were given seamless management under humanistic care. The specific operation methods are as follows:

Humanistic nursing: (1) Nursing staff need to do a good job in interviewing patients, know the patient's condition through effective communication and make the evaluation, and communicate with relevant doctors; according to the patient's operation method, condition and anesthesia method, different methods are adopted to manage patients; at the same time, formulate a set of personalized precautions for patients. Since ICU patients are relatively cumbersome and large-scale, both anesthesia and surgical procedures are more complicated. Therefore, nurses need to nurse patients carefully and cautiously. (2) Nursing staff need to adjust the temperature in the ward, preferably between 22 and 25°C, and the humidity is about 55% to ensure the comfort; do a good job in keeping the patient warm, observe the blood flow in the compressed part, and record the patient's vital signs in real time; (3)nursing staff should ensure that the drainage tube is fixed securely, the infusion channel unobstructed and the exposed parts covered; after the ward, the nursing staff make handover and communication with the nursing staff in the ward, including intraoperative responses and blood loss, etc., comfort the family members, and give them spiritual encouragement.

Seamless management: (1) Preparation stage is going to establish a seamless management team to cooperate in the entire

nursing process, with the deputy chief nurse or the chief nurse as the team leader responsible for management guidance and supervision. The team consists of 6 nurses. They are responsible for completing the basic nursing operations with due diligence. At the same time, they put forward corresponding suggestions according to the problems existing in the nursing process. The goal of the seamless management team is to reduce nosocomial infections. (2) Contents: ① nutritional support management: since ICU patients are relatively weak, their immunity is low, and their defense and resistance to foreign bacterial infection is weak so the first thing to do is to strengthen nutritional support for patients, which is also the most important thing to do as one of the basic nursing cares. Nursing staff provide nutritional support to patients in the following ways of parenteral support, on-site nutrition, etc., but it should be noted that in the process of nutrient solution deployment they should be prepared and used immediately. In theory, it should not exceed 24 h to avoid the deterioration of the nutrient solution and cause unnecessary impact on the patient, and the container for placing the nutrient solution should be dry and sterile; ② abdominal infection management: generally speaking, the occurrence of abdominal infection. It is usually due to self-infection so the drug susceptibility test results should be tested for the patient in time, and appropriate sensitive antibiotics should be selected for the patient, and antibiotics should be used correctly if the patient has related symptoms;③respiratory tract infection management: intraoral care: guide the patient to take half in the sitting position, use chlorhexidine solution to rinse the patient's mouth every day to ensure the cleanliness of the oral cavity; environmental management: it is necessary to purify the ambient air regularly, mainly using a turbocharger to ensure that the temperature in the ward is maintained at 23∼25°, the humidity is about 55% in order to ensure that the patient's airway will not be excessively dry; a humid heat exchanger can be used to properly humidify it; ventilator management: it is necessary to place the humidifier sterile liquid and thread. The tube should be replaced regularly, and the condensed water should be cleaned up in time to avoid reverse flow; ④ urinary tract infection management: after the patient is indwelled by the catheter, the vulva and urethral orifice should be disinfected with iodine with a concentration of 0.05%. The frequency is 3 times per day. After the patient defaces, the patient's perineum should also be cleaned, and the patient's urethral opening should be scrubbed at the same time in order to shorten the indwelling time of the patient's urinary catheter as much as possible to avoid complications related to the urinary catheter. ⑤ skin and soft tissue care: for ICU patients, there will be surgical incisions and puncture sites, and the skin at this site is relatively fragile so it needs to be disinfected daily, and at the same time, use a one-time sterile applicator and require frequent replacement, closely observe the skin around the puncture site of the patient; if there is redness, pain or fever, the tissue should be taken immediately for bacterial culture to ensure that the infection of the patient can be controlled in time. For long-term bedridden patients, nursing staff should instruct patients to turn over regularly to avoid local pressure and pressure ulcers.

Observation of indicators

Indicator observation includes the following steps:

(1) The nursing quality of the two sets was compared, and the nursing quality survey scale prepared by the hospital was used to evaluate the nursing quality. The Cronbach's alpha value of this scale is 0.787.

(2) The blood infection rate of the two sets of patients was compared, and the blood infection rate (%) = the number of blood infections / the total number of people × 100%.

(3) The anxiety and depression of the two sets of patients before and after nursing were compared. The self-rating anxiety scale (SAS) was used to determine the degree of anxiety. The scale has a total of 20 items, and the scoring standard for each item is 1–4 points. A score >50 indicates that the patient has an anxiety tendency, and the higher the score indicates a higher level of anxiety in the patient (20). The self-rating depression scale (SDS) is used to assess the degree of depression. The scale has 20 items, and the scoring standard for each item is 1 to 4 points. A score >53 points indicates that the patient is prone to depression, and the higher score indicates that the degree of depression is more severe (21).

(4) To compare the nursing satisfaction of the two sets of patients, the nursing satisfaction survey scale prepared by the hospital was used for evaluation, including 10 items such as work attitude, nursing effect, nursing technology, etc. Each item was

TABLE 1 Contrast of nursing quality between two sets ($\bar{x} \pm s$, points).

Set	Theoretical nursing quality	Quality of practice nursing	Integrated nursing quality
Control set ($n = 55$)	75.25 ± 5.39	72.52 ± 5.71	73.58 ± 5.26
Study set ($n = 55$)	95.33 ± 4.16	94.16 ± 5.27	94.89 ± 4.26
t	−21.827	−20.654	−23.349
P	<0.001	<0.001	<0.001

TABLE 2 Contrast of blood infection rates between the two sets of patients.

Set	Risk of blood infection (%)
Control set ($n = 55$)	14 (25.45)
Study set ($n = 55$)	3 (5.45)
$\chi 2$	8.419
P	0.004

scored as 0–10 points <95 points are satisfied, 60-<80 points are fair, and <60 points are unsatisfactory. Satisfaction = very satisfied rate + satisfaction rate. The Cronbach's alpha value of this scale is 0.763.

(5) To compare the difference in APACHE II score between the two sets, using the APACHE II scoring system (22), the vital signs, blood routine, liver and kidney function, electrolytes, blood gas analysis and Glasgow coma score within the first 24 h of ICU admission were collected (23). The total data APS (A), age adjustment (B), chronic health adjustment (C) inspection data were obtained, and all data were entered into the APACHE II scoring scale, and finally A + B + C was obtained to obtain the APACHE II score. The patients were divided into sets according to the APACHE II total score: severe set, with scores ≤15 points; critically ill set, between 16 and 35 points; extremely critically ill set, >35 points.

(6) The correlation between nursing quality score, blood infection and APACHE II score of patients was compared.

Statistical methods

In this study, all the data were organized, and a corresponding database was established for it. All databases were entered into SPSS 26.0 for data processing, and the measurement data was tested for normality, expressed as ($\bar{x} \pm s$), and between sets as t-test; the rate is expressed as %, and the test method is $\chi 2$; the correlation between nursing quality, blood infection risk and APACHE II score was analyzed by Pearson; when $P < 0.05$, the difference between the data was statistically large.

Experimental results and analysis

Contrast of nursing quality between the two sets

The scores of nursing quality in the study set, including theory, practices and comprehensiveness were notoriously higher than those in the control set (all $P < 0.05$), as shown in Table 1.

Contrast of blood infection rates between the two sets of patients

The risk rate of blood infection in the study set during hospitalization was 5.45%, which was notoriously lower than

TABLE 3 Contrast of anxiety and depression before and after nursing between the two sets of patients ($\bar{x} \pm s$, points).

Set	SAS		SDS	
	Before nursing	After nursing	Before nursing	After nursing
Control set ($n = 55$)	78.53 ± 4.36	69.36 ± 3.89*	79.85 ± 5.23	65.16 ± 4.23*
Study set ($n = 55$)	77.59 ± 4.45	36.23 ± 3.69*	80.25 ± 5.39	33.15 ± 3.19*
t	1.119	45.824	−0.395	44.808
P	0.266	<0.001	0.694	<0.001

Contrast with the same set before nursing, *P < 0.05.

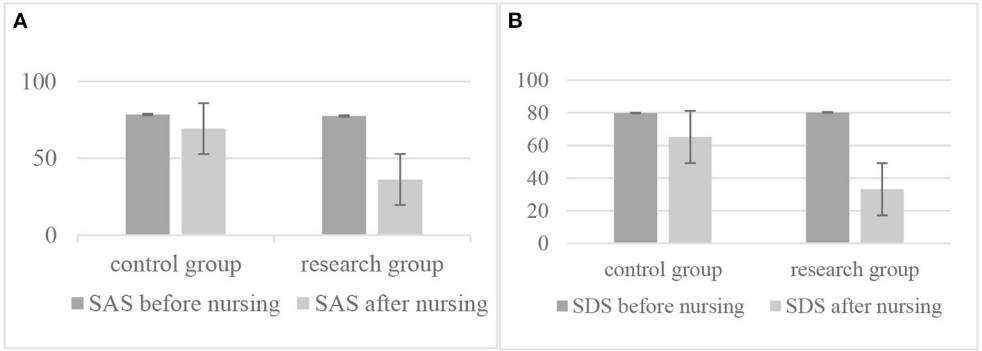


FIGURE 1
Contrast of anxiety and depression in the two sets of patients before and after nursing: **(A)** The contrast of anxiety scale between control group and research group before and/or after nursing; **(B)** The contrast of depression scale between control group and research group before and/or after nursing.

TABLE 4 Contrast of nursing satisfaction of two sets of patients.

Set	Very satisfied	Satisfy	Generally	Dissatisfied	Satisfaction
Control set ($n = 55$)	14 (25.45)	21 (38.18)	5 (9.09)	15 (27.27)	35 (63.64)
Study set ($n = 55$)	30 (54.55)	10 (18.18)	9 (16.36)	6 (10.91)	49 (89.09)
$\chi 2$					9.872
P					0.002

that in the control set, which was 25.45% ($P < 0.05$), as shown in Table 2.

Contrast of anxiety and depression in the two sets of patients before and after nursing

Before nursing, there was no large difference in the SAS and SDS scores between the two sets ($P > 0.05$). After nursing, the SAS and SDS scores of the two sets are decreased, and the study set is lower than the control set ($P < 0.05$), as shown in Table 3 and Figure 1.

Contrast of nursing satisfaction of two sets of patients

The total satisfaction with nursing in the study set was 89.09%, which was notoriously higher than that in the control set (63.64%) ($P < 0.05$), as shown in Table 4.

Differences in APACHE II scores before and after nursing between the two sets of patients

Before nursing, the APACHE II scores of the two sets are higher, and the difference was not statistically large ($P > 0.05$), as shown in Table 5.

TABLE 5 Contrast of APACHE II scores before and after nursing between the two sets of patients ($\bar{x} \pm s$, points).

Set	Before nursing	After nursing
Control set ($n = 55$)	33.25 ± 5.37	$22.47 \pm 4.58^*$
Study set ($n = 55$)	32.93 ± 5.54	$14.85 \pm 2.39^*$
t	0.308	10.939
P	0.759	<0.001

Contrast with before nursing, *means $P < 0.05$.

TABLE 6 Correlation between various scores of nursing quality and APACHE II score.

Care quality	r	P
Theoretical nursing quality	−0.973	<0.001
Quality of practice nursing	−0.945	<0.001
Comprehensive quality of care	−0.972	<0.001

Correlation between nursing quality score, blood infection risk and APACHE II score

Correlation analysis between scores of nursing quality and APACHE II score

There was a large negative correlation between various scores of nursing quality and APACHE II score ($P < 0.05$), as shown in Table 6 and Figure 2.

Correlation between blood infection risk and APACHE II score

There was a large negative correlation between the risk of blood infection and the APACHE II score ($P < 0.05$), as shown in Table 7.

Conclusion and the future work

For neurological ICU patients, the seamless nursing model of humanistic care can improve their nursing quality reduce the risk of blood infection and APACHE II score, promote patient recovery, and improve patient prognosis, which is worthy of clinical application. Although this paper has achieved certain results, it still has certain limitations. Since this study did not evaluate the prognosis of patients, it is impossible to determine the long-term effect and compliance of nursing care. Therefore, in future research, the scope of sample selection should be expanded and extend the follow-up stage and further evaluate the long-term effect of the seamless nursing model of humanistic care on neurosurgical ICU patients.

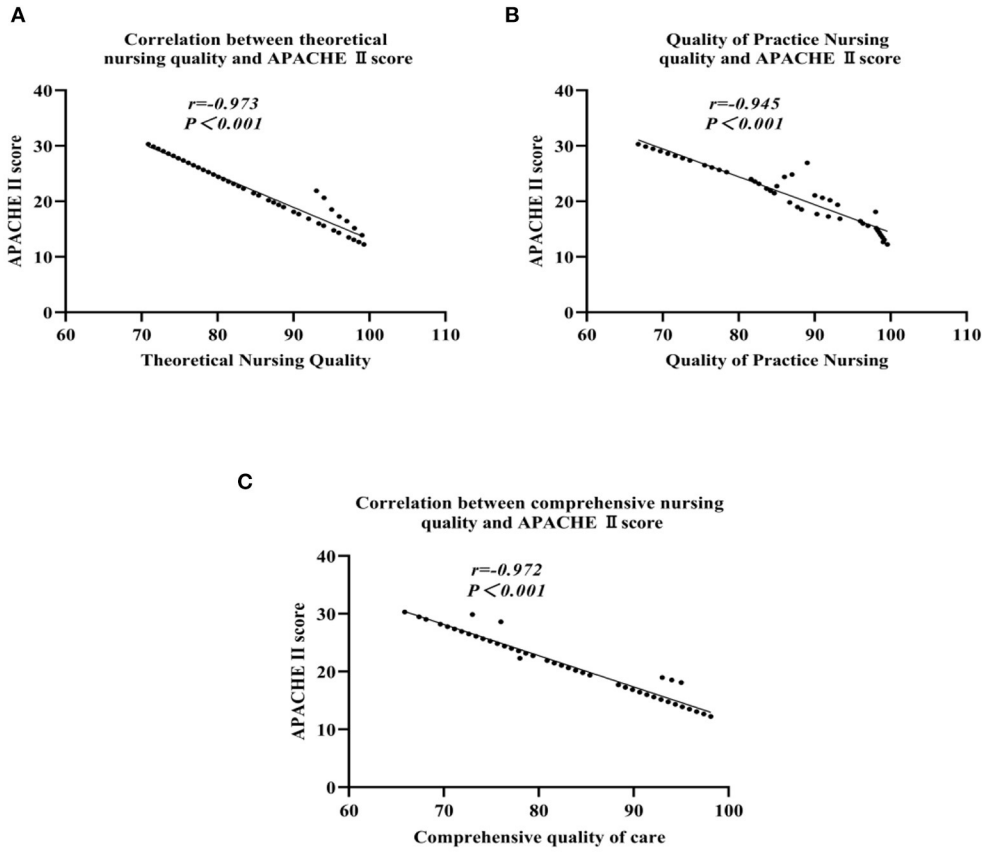


FIGURE 2
Correlation analysis of various scores of nursing quality and APACHE II score: **(A)** The theoretical nursing quality; **(B)** The practice nursing quality; **(C)** The comprehensive nursing quality.

TABLE 7 Correlation between blood infection and APACHE II score.

Index	*r*	*P*
Risk of blood infection	−0.652	<0.001

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Ethics statement

The studies involving human participants were reviewed and approved by Zhejiang Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Zhao H, Zhang J, LuY, Jin J. Neoadjuvant chemotherapy in combination with surgery in the treatment of local advanced breast cancer. *Pak J Med Sci.* (2019) 35:1402. doi: 10.12669/pjms.35.5.310

2. Algahtani R, Merenda A. Multimorbidity and critical care neurosurgery: minimizing major perioperative cardiopulmonary complications. *Neurocrit Care.* (2020) 15:833. doi: 10.1007/s12028-020-01072-5

3. Wang Q, Wei L. Application of seamless nursing management in emergency-icu patient safety transfer and handover. *J Adv Med Sci.* (2019) 2:4. doi: 10.30564/jams.v2i2.514

4. Li Y, Xu F. Application effect of detail nursing in neurosurgery patients. *Clin Nurs Res.* (2021) 5:31–9. doi: 10.26689/jcnr.v5i2.1895

5. Kim M, Lee S, Park W, Ahn JC. 1722: anaphylactic shock followed by indocyanine green videoangiography in neurosurgery and ICU care. *Crit Care Med.* (2020) 48:725–9. doi: 10.1097/01.ccm.0000649188.19735.8f

6. Al-Ma'Aitah M, Alwadain A, Saad A. Transmission adaptive mode selection (TAMS)method for internet of things device energy management. *Peer Peer Netw Appl.* (2021) 14: 366–9. doi: 10.1007/s12083-020-00937-y

7. Kang LN. Application effect of seamless nursing in elderly patients with coronary heart disease. *Smart Healthcare.* (2019) 3:331–5. doi: 10.19335/j.cnki.2096-1219.2019.28.059

8. EM O, Fafolahan, Pierre S. A seamless mobility management protocol in 5G locator identificator split dense small cells. *IEEE Trans Mob Comp.* (2020) 19:1745–59. doi: 10.1109/TMC.2019.2915071

9. Parasuraman S, Rajaraman G. Optimized seamless handover scheme with information and decision server "resisted partially distributed mobility management in heterogeneous 5th generation wireless networks. *Int J Commun Syst.* (2021) 7:96. doi: 10.1002/dac.4795

10. Guo X, Neurosurgery D O. Observation on the effect of integrated nursing care after cerebral glioma surgery. *Chin Contin Med Educ.* (2019) 27:41–5. doi: 10.3969/j.issn.1674-9308.2019.14.074

11. Su X, He S, Wu Z, Liu Q, Li S, Liu X. Observation on the application effect of clinical nursing pathway table in ICU patients with mechanical ventilation. *Nursing Practice and Research.* (2019) 6:45–9.

12. Romero-Garcia R, Erez Y, Oliver G, Owen M, Merali S, Poologaindran A, et al. Practical application of networks in neurosurgery: combined 3-Dimensional printing, neuronavigation, and preoperative surgical planning-sciencedirect. *World Neurosurg.* (2020) 137:226–8. doi: 10.1016/j.wneu.2020.01.085

13. Kalra D, Pradhan MR. Enduring data analytics for reliable data management in handling smart city services. *Soft Comput.* (2021) 1–13. doi: 10.1007/s00500-021-05892-1

14. Huo RL, Lai H, Lin Y, et al. Application of the "Rain Classroom" hybrid teaching mode in undergraduate nursing students in neurosurgery. *J Snake.* (2019) 6:524–8. doi: 10.3969/j.issn.1001-5639.2019.01.067

15. Neto EP, Silva F, Schneider LM, Neto AV, Immich R. Seamless MANO of multi-vendor SDN controllers across federated multi-domains. *Comput Netw.* (2020) 186:107752. doi: 10.1016/j.comnet.2020.107752

16. Liu W, Shi J. Application effect of humanized nursing in hemodialysis for diabetic nephropathy uremia. *Clin Res Pract.* (2019) 27:587–9. doi: 10.19347/j.cnki.2096-1413.201926075

17. Xiao S, Hospital PP. Application effect of standardized propaganda process and video nursing in neurosurgery health education. *Med Innov Chin.* (2019) 7:352–9.

18. Li H. Effects observation of targeted nursing on lung infection patients in NICU. *Chin Contin Med Edu.* (2019) 21:62–7. doi: 10.3969/j.issn.1674-9308.2019.22.081

19. Liang W, Deng B, Luo W, et al. Application effect of continuous humanistic care concept in the nursing of ICU patients. *Clin Med Eng.* (2019) 33:1253–6. doi: 10.3969/j.issn.1674-4659.2019.06.0845

20. Wang Zheyun Y, Yan W, et al. Research on the formulation and application of enhanced recovery airway management plan for cardiac surgery patients. *Chin J Nurs.* (2020) 55:1605–11. doi: 10.3761/j.issn.0254-1769.2020.11.001

21. Zhang T, Yu GD, Gu P, Tang Q, Jin Y, He XC, et al. Evaluation and analysis of anxiety and depression status and quality of life in vasomotor rhinitis. *Chin J Otolaryngol Head Neck Surg.* (2020) 55:769–73. doi: 10.3760/cma.j.cn115330-20191230-00787

22. Weiqun H, Xiaoqing L, Yimin L, et al. Analysis of risk factors for bloodstream infection caused by pulmonary-derived Acinetobacter baumannii in mechanically ventilated patients: a 5-year Observational Study from the Real World. *Chin Crit Care Emerg Med.* (2016) 28:487–91.

23. Ji Y, Li L, Xindi J. Contrast of the evaluation value of the global unresponsiveness scale and the glasgow coma scale in the assessment of disturbance of consciousness and short-term prognosis in patients in neurointensive care units. *PLA Nurs J.* (2019) 36:18–21.

frontiers | Frontiers in Public Health

ORIGINAL RESEARCH
published: 09 September 2022
doi: 10.3389/fpubh.2022.914973

Retinal Vessel Segmentation Based on B-COSFIRE Filters in Fundus Images

Wenjing Li[1], *Yalong Xiao*[2,3], *Hangyu Hu*[4], *Chengzhang Zhu*[2,3]*, *Han Wang*[3,5], *Zixi Liu*[3,5] *and Arun Kumar Sangaiah*[6]

[1] Hunan Province People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha, China, [2] The College of Literature and Journalism, Central South University, Changsha, China, [3] Mobile Health Ministry of Education-China Mobile Joint Laboratory, Changsha, China, [4] School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China, [5] School of Computer Science and Engineering, Central South University, Changsha, China, [6] School of Computing Science and Engineering, Vellore Institute of Technology University, Chennai, Tamil Nadu, India

Retinal vessel extraction plays an important role in the diagnosis of several medical pathologies, such as diabetic retinopathy and glaucoma. In this article, we propose an efficient method based on a B-COSFIRE filter to tackle two challenging problems in fundus vessel segmentation: (i) difficulties in improving segmentation performance and time efficiency together and (ii) difficulties in distinguishing the thin vessel from the vessel-like noise. In the proposed method, first, we used contrast limited adaptive histogram equalization (CLAHE) for contrast enhancement, then excerpted region of interest (ROI) by thresholding the luminosity plane of the CIELab version of the original RGB image. We employed a set of B-COSFIRE filters to detect vessels and morphological filters to remove noise. Binary thresholding was used for vessel segmentation. Finally, a post-processing method based on connected domains was used to eliminate unconnected non-vessel pixels and to obtain the final vessel image. Based on the binary vessel map obtained, we attempt to evaluate the performance of the proposed algorithm on three publicly available databases (DRIVE, STARE, and CHASEDB1) of manually labeled images. The proposed method requires little processing time (around 12 s for each image) and results in the average accuracy, sensitivity, and specificity of 0.9604, 0.7339, and 0.9847 for the DRIVE database, and 0.9558, 0.8003, and 0.9705 for the STARE database, respectively. The results demonstrate that the proposed method has potential for use in computer-aided diagnosis.

Keywords: retinal vessel segmentation, COSFIRE, postprocess, computer-aided diagnosis, medical image segmentation

OPEN ACCESS

Edited by:
Celestine Iwendi,
School of Creative Technologies
University of Bolton, United Kingdom

Reviewed by:
Ebuka Ibeke,
Robert Gordon University,
United Kingdom
Shweta Agrawal,
Sage University, India

***Correspondence:**
Chengzhang Zhu
anandawork@126.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *07 April 2022*
Accepted: *14 June 2022*
Published: *09 September 2022*

Citation:
Li W, Xiao Y, Hu H, Zhu C, Wang H, Liu Z and Sangaiah AK (2022) Retinal Vessel Segmentation Based on B-COSFIRE Filters in Fundus Images. Front. Public Health 10:914973. doi: 10.3389/fpubh.2022.914973

INTRODUCTION

The color fundus image (**Figure 1A**) is a non-invasive tool generally used to diagnose various pathologies, including diabetic retinopathy, glaucoma, and age-related macular degeneration. Retinal vessels are the only part of the blood circulation system of humans that can be non-invasively observed directly (1, 2). Inspection of the attributes of the retinal vessel such as width, tortuosity, branching pattern, and angles can play significant roles in early disease diagnosis. Moreover, vessel segmentation is a key and indispensable processing step for further

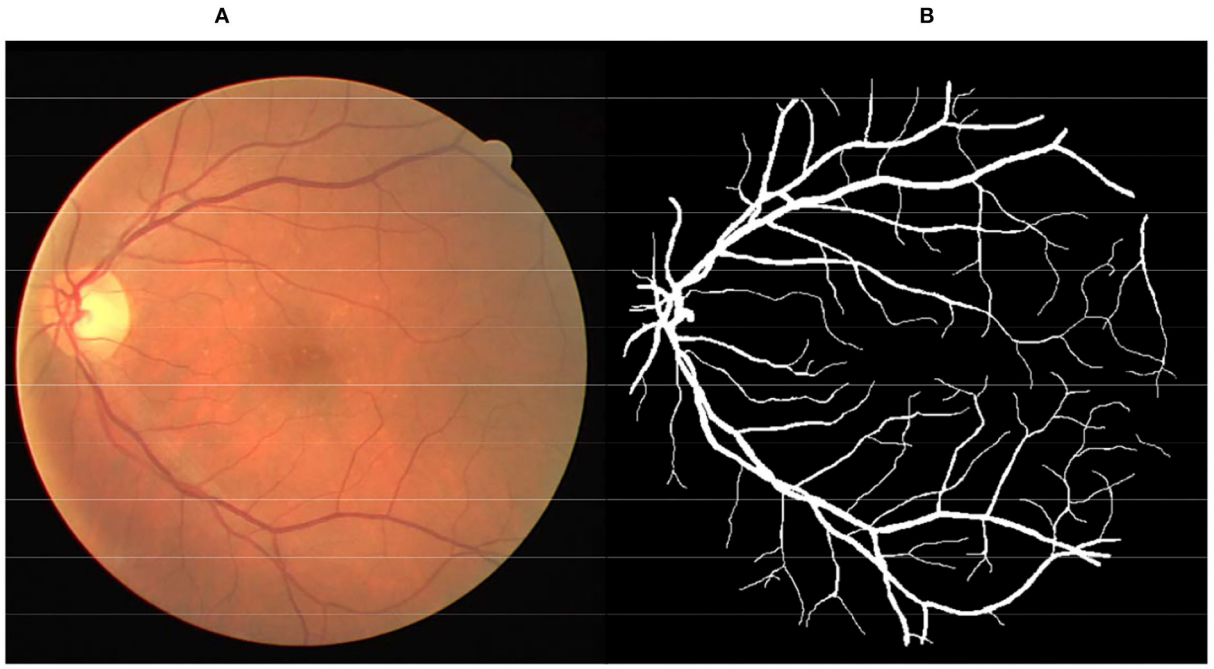


FIGURE 1 | (A) Color fundus image from DRIVE database. **(B)** Corresponding manual segmented image.

processing operations. This means that the accuracy of vessel segmentation greatly affects the diagnostic effect. But, it is a work of redundancy and boredom for professional ophthalmologists, as shown in **Figure 1B**. Consequently, to save medical resources and improve the effect of diagnosis, the effective vessel segmentation algorithm with high accuracy and less time consumption is remarkably necessary in computer-aided diagnosis.

A bar-selective COSFIRE filter (B-COSFIRE) is highly efficient for bar-shaped structure detection (3). Therefore, it is a novel and valid method for the automatic segmentation of blood vessels. It is based on the Combination of Receptive Field (CORF) computational model of a simple cell in visual cortex (4) and its implementation, called Combination of Shifted Filter Response (COSFIRE) (5). The COSFIRE filter is trainable, meaning it is not predefined in the implementation, but it is determined from a user-specified prototype pattern (3). COSFIRE has the versatility that can be configured to solve many image processing and pattern matching tasks.

In this article, we propose an efficient retinal vessel segmentation method based on the B-COSFIRE filter based on a previous work mentioned in Azzopardi et al. (3). The original method has good time complexity, but its performance is not as satisfactory as we expected. In the proposed method, we have made many attempts, such as employing denoise operation and post-processing operation. The proposed method has achieved a better performance result without losing the nature of good time complexity.

The rest of this article is organized as follows: In Related Works, we examine the existing blood vessel segmentation method and other applications of COSFIRE filters. In Proposed Methodology, we present our method explicitly. In Experimental Results and Analysis, we introduce the datasets used for experiments and reported the experimental results and the comparisons with existing methods. The discussion for future work and conclusion are presented in Findings and Conclusion, respectively.

RELATED WORKS

Due to the importance of vascular segmentation in computer-aided diagnosis, an efficient vessel segmentation algorithm has always been a research hot spot. Existing approaches for retinal blood vessel segmentation in fundus images can be divided into two groups: unsupervised methods, which include matched filtering, vessel tracking, model-based approaches, and morphological processing, and supervised methods, which use feature vectors to train a binary classification model (6–8). Whether the method is supervised or not depends only on whether manual marking information with a priori is used.

Supervised methods use pixel-wise feature vectors with labeled information, where manually segmented images are referred to as the gold standard, to train a classifier that can distinguish between vascular and non-vascular pixels. This kind of method mainly includes the following two steps: form pixel-wise feature vectors by feature extraction methods and learn

a classification model based on vessel and non-vessel training feature vectors (9). Soares et al. (10) reported a segmentation method based on a Bayesian classifier combined with multi-scale analysis for two-dimensional Gabor wavelet transform. Marin et al. (11) proposed a supervised method for retinal vessel segmentation, by applying a multi-layer neural network to classify pixels based on moment-invariant features. Fraz et al. (12) proposed a classification scheme that fuses boosted and bagged decision trees. Aslani et al. (13) presented a supervised method using the random forest classifier, which constructs a rich collection of a 17-dimensional feature vector including B-COSFIRE filter response and trains a random forest classifier to accomplish the segmentation of fundus images. Strisciuglio et al. (14) proposed a method of retinal vessel segmentation by transforming and re-scaling the features composed of the responses of the bank of selected B-COSFIRE filters to train a support vector machine classifier. Zhu et al. (15) presented a supervised ensemble method for segmenting the retinal vessels by extracting a 39-dimensional feature vector to train the extreme learning machine classifier.

Unsupervised methods mainly extract pathological features through linear operation filtering techniques with predefined kernels. Al-Diri et al. (16) proposed an unsupervised method based on the active contour model to achieve retinal vessel measurement and segmentation. Lam et al. (17) presented a multi-concavity modeling approach based on differentiable concavity measure, which can simultaneously process retinal images of two different health states. Azzopardi et al. (3) proposed a method of retinal vessel segmentation based on the COSFIRE approach, called B-COSFIRE filter, by constructing two kinds of the B-COSFIRE filter which are selective for vessel and vessel-ending, respectively. Khan et al. (18) presented an unsupervised method of vasculature segmentation, by applying pixel AND operation between the vessel location map and B-COSFIRE segmentation image. Bahadarkhan et al. (19) proposed a less computational unsupervised automated technique with promising results for the detection of retinal vasculature by using morphological Hessian-based approach and region-based Otsu thresholding. Khan et al. (20) used distinctive preprocessing steps, thresholding techniques, and post-processing steps to enhance and segment the retinal blood vessels. Khan et al. (21) proposed a framework with fast execution and competing outcomes using MISODATA and B-COSFIRE filters to produce better segmentation results.

Generally, supervised methods perform better than other kinds of methods in retinal vessel segmentation, but supervised methods consume a lot of time in the process of classifier training. On the other hand, the advantage of unsupervised learning is that it does not require artificial label information to train the classifier, which makes it more versatile and easier to implement (6, 22). To sum up, a robust method that can have good performance results and does not require much preparation time is necessary for computer-aided diagnosis.

Deep learning has always been a hot topic in the field of computer research. Deep learning can achieve high-quality performance but requires more training time than supervised learning (23, 24). The advantages and disadvantages of deep learning are obvious. Many scholars have attempted using methods of deep learning to segment vessels in retinal images. Melinsca et al. (25) proposed a method using deep neural network to segment retinal vessels. Khalaf et al. (26) proposed a method of retinal vessel segmentation based on convolutional neural networks for deep feature learning. Fu et al. (27) employed deep learning network and fully connected conditional random fields to accomplish retinal vessel segmentation. Deep learning can often achieve better performance than machine learning; however, deep learning requires much training time and better hardware equipment conditions. Yue et al. (28) introduced the multi-scale input layer and dense block to the conventional U-net so that the network can make use of richer spatial context information. Cheng et al. (29) added a dense block to the U-Net network to make each layer's input come from all the previous layer's output, thereby improving the segmentation accuracy of small blood vessels. In order to follow the discussion easily, we summarize the discussed studies in **Table 1**.

The COSFIRE method was proposed by Azzopardi et al. (5), based on the CORF computational model. Due to the fast and accurate nature of COSFIRE, it has received great attention in the field of image processing and pattern recognition. In addition to the application of retinal vascular segmentation as mentioned before, there are some applications based on the COSFIRE filter in other directions. Azzopardi et al. (33) employed descriptors of different shapes based on trainable COSFIRE filters to recognize handwritten digits, detect vascular bifurcations in segmented retinal images in Azzopardi et al. (34), and achieve gender recognition from face images in Azzopardi et al. (35). Gecer et al. (36) proposed a method that can recognize objects with the same shape but different colors, by configuring different COSFIRE filters in different color channels. Guo et al. (37, 38) further developed the COSFIRE method, by configuring the COSFIRE filter with the inhibition mechanism to recognize architectural and electrical symbols and to detect key points and recognize objects.

The aforementioned work has the following problems:

1) Related supervision methods (10–15) have high computational cost and high time cost in the process of model training. Supervised methods need to use a large number of pixel-level data labels for supervised computation, and it consumes much time to produce segmentation results. As shown in **Tables 6**, **7**, the overhead of the relevant supervised methods in the segmentation time is about 1 min.
2) When considering the combined effects of performance and time overhead, some unsupervised (16–21, 30) methods lack efficient application value. Some unsupervised methods have higher time overhead when obtaining higher segmentation performance. While another part of the unsupervised method achieves lower time overhead, the segmentation performance is also reduced. That is to say, these methods cannot guarantee lower time overhead and a better segmentation effect at the same time; in addition, these methods are not versatile and cannot achieve good results in multiple dataset tests.

TABLE 1 | Related work of retinal segmentation methods.

	Reference	Year	Segmentation methods	Technology
Supervised methods	Soares et al. (10)	2006	Combining two-dimensional Gabor wavelet transform multi-scale analysis and Bayesian classifier	2-D Gabor wavelet, bayesian classifier with Gaussian mixtures
Supervised methods	Marin et al. (11)	2011	a multi-layer neural network to classify pixels based on moment-invariant features	A multi-layer feedforward network
Supervised methods	Fra et al. (12)	2012	a classification scheme based on an ensemble of boosted and bagged decision trees	Boosted and bagged decision trees
Supervised methods	Aslani and Sarnel (13)	2016	using the random forest classifier based on a hybrid feature vector for pixel characterization	Random forest classifier, morphological top-hat, B-COSFIRE filter, multi-scale Gabor wavelet
Supervised methods	Strisciuglio et al. (14)	2016	transforming and re-scaling the features composed of the responses of the bank of selected B-COSFIRE filters to train support vector machine classifier	B-COSFIRE filters, SVM classifier, GMLVQ, Genetic algorithm
Supervised methods	Zhu et al. (15)	2016	extracting a 39-dimensional feature vector to train the extreme learning machine classifier	Classification and regression tree (CART)
Unsupervised methods	Al-Diri et al. (16)	2009	An active contour model to achieve retinal vessel measurement and segmentation.	Active contour model, growing algorithm, junction resolution algorithm
Unsupervised methods	Liu and Sun (17)	1993	a multi-concavity modeling approach with differentiable concavity measure	Adaptive tracking algorithm
Unsupervised methods	Azzopardi et al. (3)	2015	proposing B-COSFIRE approach for retinal vessel segmentation	B-COSFIRE filters, CLAHE, Masking
Unsupervised methods	Khan et al. (18)	2016	a vasculature segmentation method by applying pixel AND operation between vessel location map and B-COSFIRE segmentation image	CLAHE, morphological filters, difference image of low pass filter, adaptive thresholding, VLM, B-COSFIRE filters
Unsupervised methods	Bahadarkhan et al. (19)	2016	Using morphological hessian-based approach and region-based Otsu thresholding	CLAHE, morphological filters, Hessian matrix and eigenvalues transformation, Otsu thresholding
Unsupervised methods	Khan et al. (20)	2016	proposing Modified Iterative Self Organizing Data Analysis Technique (MISODATA) for vessel segmentation	CLAHE, MISODATA algorithm, postprocessing
Unsupervised methods	Khan et al. (21)	2020	proposing a framework applicating MISODATA and B-COSFIRE filters with fast execution and competing outcomes.	CLAHE, MISODATA algorithm, B-COSFIRE filters
Unsupervised methods	Ooi et al. (30)	2021	Applying edge detection technology based on Canny algorithm.	CLAHE, Canny algorithm
Deep learning methods	Melinsca et al. (25)	2015	using deep neural network to segment retinal vessels	Deep max-pooling convolutional neural networks
Deep learning methods	Khalaf et al. (26)	2016	Using convolutional neural networks for deep feature learning	Deep Convolutional Neural Networks
Deep learning methods	Fu et al. (27)	2016	employing deep learning network and fully- connected conditional random fields to accomplish retinal vessel segmentation.	FCN, Conditional Random Fields (CRFs)
Deep learning methods	Yue et al. (28)	2019	Improved U-net with Multi-scale input layer and dense block introduced	U-net
Deep learning methods	Cheng et al. (29)	2020	Adding dense block to U-Net network	U-net
Deep learning methods	Li et al. (31)	2021	A scheme based on the combination of U-Net and Dense-Net is proposed	U-net, Dense-net, CLAHE
Deep learning methods	Feng et al. (32)	2022	encoder-decoder structure	Inception, Multiple pyramid pooling modules

3) The images that have been processed by binary thresholding are destroyed by noise. Distinguishing the thin vessel from the vessel-like noise is still a challenge (18, 20, 21).

In this article, an improved unsupervised method based on the B-COSFIRE filter is presented. Compared with traditional methods, the proposed method has stronger robustness. The proposed unsupervised method uses image contrast enhancement algorithms and morphological operations and uses the B-COSFIRE filter, which can effectively extract bar-like blood vessels in the fundus image. While improving the effect, it reduces the time overhead. At the same time, our method uses a post-processing algorithm based on connected domains, which can effectively distinguish small connected blood vessel

pixels from vessel-like noise. This proves once again that the proposed method has a great application capability in computer-aided diagnosis.

PROPOSED METHODOLOGY

Vessel segmentation is a basic step in image processing of the fundus image; thus, the process should be fast and efficient. The main idea is taking advantage of the high efficiency of the B-COSFIRE filter and employing other operations that do not need much consumption of time to obtain a better result. **Figure 2** represents a flowchart of our proposed method with main processing steps. **Algorithm 1** shows the implementation steps of our proposed system. The mission of the B-COSFIRE filter is to select all vessels from the fundus image, while the other operation is to enhance vascular features, denoise background noise, and reduce error classification.

Preprocessing

Before employing operations to segment vessels, we used the following operations to improve or enhance the characteristics of blood vessels:

a. Green channel extraction: The green band is extracted from RGB retinal images, and the FOV mask (field of view) image is obtained by thresholding the luminosity plane of the CIELab version of the original RGB image. Previous works by Niemeijer et al. (39), Mendonca et al. (40), Soares et al. (10), and Ricci et al. (41) have demonstrated that the green channel of RGB fundus images can best highlight the difference between blood vessels and background, whereas the red and blue channels show low contrast and are very noisy.
b. Mask produce: The FOV mask is an important tool for determining the ROI in the process of vessel segmentation. Although the DRIVE dataset provides the FOV mask, most other datasets do not. For the sake of versatility of the proposed method, we adopt this method to obtain FOV mask images by shifting the original RGB images to CIELab version and then thresholding the luminosity plane.
c. CLAHE: Next, the contrast-limited adaptive histogram equalization (CLAHE) algorithm is used to enhance images for expanding vessels characteristics. The CLAHE algorithm can effectively limit noise amplification in relatively uniform areas and improve local contrast, which is usually used as a preprocessing step in retinal image analysis research (42).

Segmentation Processing

The proposed segmentation method is based on the feature of the B-COSFIRE filter with the highest correlation to the bar/vessel shape. A B-COSFIRE filter was originally proposed in Azzopardi et al. (3). It takes the responses of a group of difference-of-Gaussians (DoG) filters at certain positions with respect to the center of its area of support as input. In the proposed method, the B-COSFIRE filter has been used for efficiently obtaining vessel-like structures. The B-COSFIRE filter is trainable and is configured to be selective for bar-like structures (3). The term trainable refers to the ability of determining these positions in an automatic configuration process by using a synthetic vessel or vessel-ending image.

The application of the B-COSFIRE filter consists of four main process: Convolution with DoG filters, blurring response, shifting the blurred responses, and estimating a point-wise weighted geometric mean. In the following, we will introduce these steps.

A center-on DoG function with a positive central region and a negative surround denoted by $DoG_\sigma\left(x,\ y\right)$ is given by (3):

$$DoG_\sigma(x,\ y)\overset{\text{def}}{=} \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) - \frac{1}{2\pi(0.5\sigma)^2}\exp\left(-\frac{x^2+y^2}{2(0.5\sigma)^2}\right) \quad (1)$$

where σ is the standard deviation of the Gaussian function that determines the extent of the surround, 0.5σ is the standard deviation of the inner Gaussian function, and (x, y) represents a pixel location of an image I. The response of a the *DoG* filter $C_\sigma\left(x,\ y\right)$ with a kernel function $DOG_\sigma\left(x-x\prime\ ,\ y\ -y\prime\right)$ is computed by convolution, where $\left(x\prime\ ,\ y\prime\right)$ represents intensity distribution of image I:

$$C_\sigma(x,\ y)\overset{def}{=}|I*DoG_\sigma|^+ \quad (2)$$

where $|\ \cdot\ |^+$ half-wave rectification operation is to suppress (set to 0) the negative values.

In the proposed B-COSFIRE filter (3), every point i is described by a tuple of three parameters $DoG(\sigma_i,\ \rho_i,\ \phi_i)$, where σ_i represents the standard deviation of the *DoG* filter that provides the input, ρ_i and ϕ_i represent the polar coordinates of the B-COSFIRE filter, and a set of three tuples of a B-COSFRE filter is denoted by $S=\ (\sigma_i,\ \rho_i,\ \phi_i|i=\ 1,\ ..,\ n)$, where n stands for the number of considered *DoG* response.

The blurring operation of the *DoG* filter is shown as follows. It allows for some tolerance in the position of the specific points.

$$\sigma'=\sigma_0'+\alpha\rho_i \quad (3)$$

where σ_0' and α are constants.

Each blurred *DoG* response is shifted by a distance ρ_i in the direction opposite to ϕ_i, and they meet at the support center of the B-COSFIRE filter. The blurred and shifted response of a *DoG* filter for each tuple $(\sigma_i,\ \rho_i,\ \phi_i)$ is denoted by $S_{(\sigma_i\ ,\rho_i\ ,\phi_i)}\left(x,\ y\right)$ in set *S*. The *i_th* blurred and shifted *DoG* response is defined as follows:

$$S_{\sigma_i,\rho_i,\phi_i}(x,\ y)\ =\ max_{x',y'}\{C_{\sigma_i}(x-\Delta x_i-x',y-\Delta y_i - y')G_{\sigma_i}(x',\ y')\} \quad (4)$$

where $-3\sigma'\le x',\ y'\le 3\sigma'$.

Last, the output of the B-COSFIRE filter is defined as the weighted geometric mean of all the blurred and shifted DoG response:

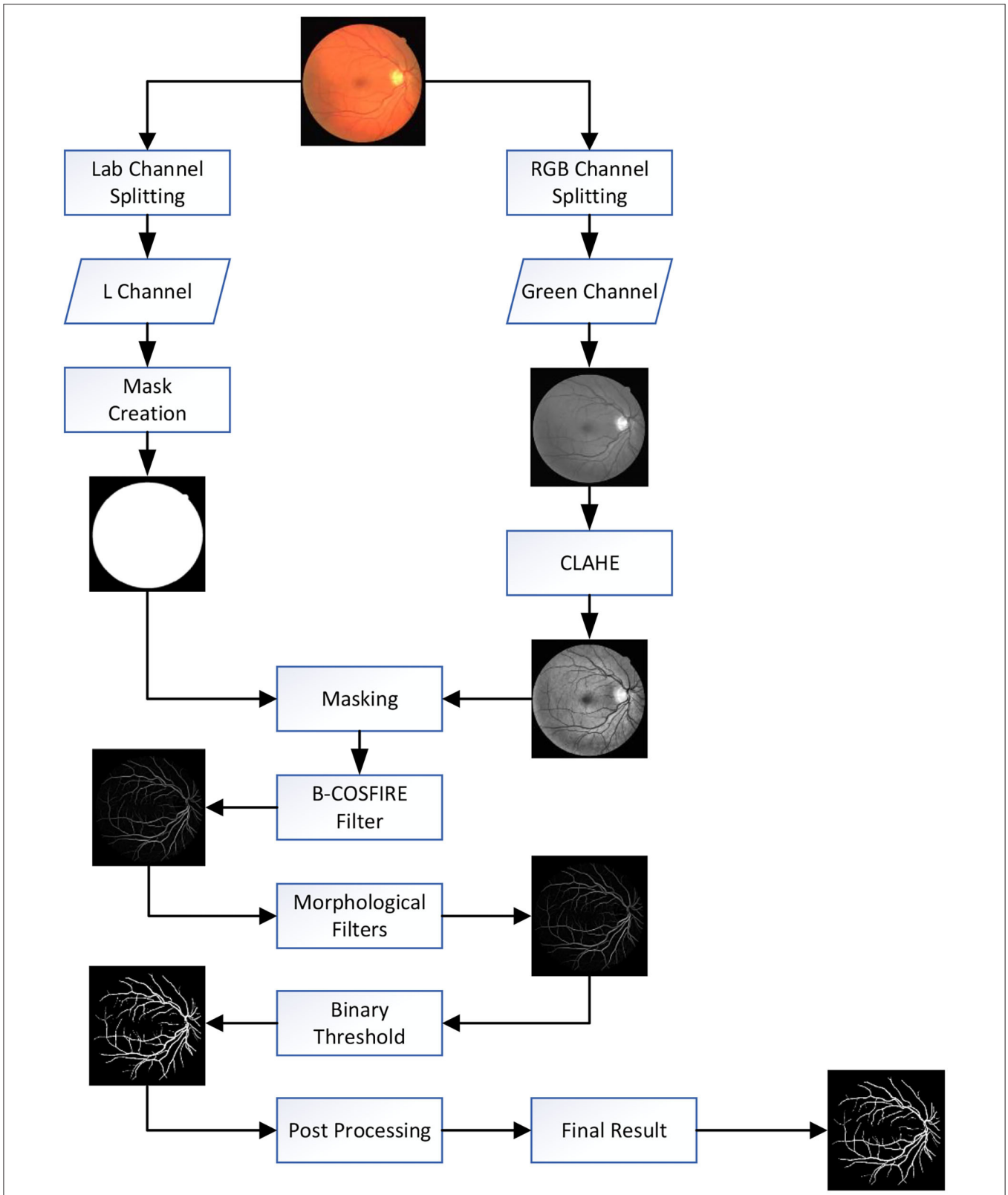


FIGURE 2 | Flowchart of proposed method: first, we using contrast limited adaptive histogram equalization (CLAHE) for contrast enhancement. Second, we threshold the luminance plane of the CIELab version of the original RGB image to produce a mask. Third, we apply the B-COSFIRE filter and morphological filter to detect blood vessels and remove noise. Fourth, the binary threshold was used for vessel segmentation. Finally, unconnected non-vessel pixels are eliminated by post-processing to obtain the final segmentation map.

Algorithm 1 | B-COSFIRE vessel segmentation algorithm.

Input: Original RGB retinal images
Output: Final segmented vessel maps
1 *Step*1:
2 Extract green band image G form Original RGB retinal images.
3 Extract CIELab version image L form Original RGB retinal images.
4 *Step*2
5 Using CLAHE algorithm to green band image G.
6 Thresholding the luminosity plane of L and produce the mask M.
7 *Step*3
8 Perform masking operation on G and M.
9 Produce the input I of the following vessel segmentation operation.
10 *Step*4
11 Applicate B-COSFIRE filter on I and produce I1.
12 *Step*5
13 Perform Morphological filters (top-hat) on I1 and produce I2.
14 *Step*6
15 Operating binary threshold on I2 and produce I3.
16 *Step*7
17 Post processing on I3.
18 Return: Final segmented vessel maps

$$r_S(x,y) \stackrel{def}{=} \left| \left(\prod_{i=1}^{|S|} \left(S_{\sigma_i,\rho_i,\phi_i}(x,y) \right)^{\omega_i} \right)^{\frac{1}{\sum_{i=1}^{|S|}\omega_i}} \right|_t \tag{5}$$

where $\omega_i = exp^{-\frac{\rho_i^2}{2\sigma^2}}$ and $|.|_t$ represent thresholding the response at a fraction $t(0 \leq t \leq 1)$. The equation is an AND-type function that a B-COSFIRE filter achieves as response when all *DoG* filter responses are >0.

Moreover, to achieve multi-orientation selectivity, the number of B-COSFIRE filters is configured by using prototype patterns in different orientations. A new set is created by manipulating the parameter ¦× of each tuple:

$$R_\psi(S) = (\sigma_i,\ \rho_i,\ \phi_i + \psi) | \forall(\sigma_i,\ \rho_i,\ \phi_i) \in S \tag{6}$$

A rotation-tolerant response is achieved by merging the response of B-COSFIRE filters with different orientation preferences and taking the maximum value at each location (x,y):

$$\hat{r}\, S(x,y) \stackrel{def}{=} \max_{\psi \in \Phi} r_{R_\psi(x,y)} \tag{7}$$

The aforementioned operation is an AND-type function that is achieved by the B-COSFIRE filter when all *DoG* filter responses are non-zero. In total, two kinds of the B-COSFIRE filter are configured (3): symmetric B-COSFIRE filter and asymmetric B-COSFIRE filter. One is selective for vessel, and the other is selective for vessel-ending. For more details, refer to Azzopardi et al. (3).

Morphological filters are used to denoise and to reduce the influence of ophthalmic disorders and to extract useful and meaningful information in small regions of images. Combining image subtraction with openings and closings result in top-hat and bottom-hat transformations (15). The role of the two kinds of transformation is the same, making non-uniform backgrounds uniform and enhancing the image contrast. The top-hat transform is used for light objects on a dark background, so it can make characteristics of vessels more apparent in the dark background. We used the top-hat transform to process the response image produced by the B-COSFIRE filter to enhance vessel structures and reduce noise. As a result, during the threshold segmentation operation, more vessel pixels will be correctly classified and noise pixels will decrease.

The top-hat transformation is defined as follows:

$$G = I - I \circ S \tag{8}$$

where I is a gray-scale image, S is a linear structuring element, and $\circ$ is opening operation. The top-hat transformation of I is defined as I minus the opening of I. Opening operation can effectively extract useful information of the background that is the same size of the structuring element so that employing top-hat transformation can relatively obtain uniform front-view information.

In this study, the structuring element is square. We experimentally found that the use of morphological top-hat transformation has less noise and better performance in vessel segmentation.

There are two ways to select threshold processing vessel response images. The first method is selecting a manual threshold for each dataset, as in Azzopardi et al. (3). This method does not need too much processing time, but the result is relatively bad. The second one is called adaptive thresholding, which automatically selects the threshold value for each image, instead of the whole dataset, as in Khan et al. (18). In proposed method, we choose the first method to select threshold standing in the point of quickly and effectively segmenting vessels. There are no significant differences between the results obtained by different methods, but the first method can save as much processing time as we expected.

Post-processing

The images that have been processed by binary thresholding are destroyed by noise. The present results of vessels segmented are far from satisfactory, that is, some vessel pixels are wrongly classified and disappear in the segmentation images. At the same time, many noise points are classified into vessels. So, we consider the post-processing measures to reduce this phenomenon to obtain a better segmentation result.

The adopted post-processing method is based on the connected domain to recover vessels. Specific steps are as follows:

*Step*1. Carry out thinning operation on binary image I_{re} and then the expansion operation with the template of 3 by 3, followed by morphology complement operation so that 0 surround by 1 in the eight neighborhoods is set to 0.

*Step*2. Repeat *Step* 1 to obtain a relatively complete vascular connected domain and the resulting image I_{fin}.
*Step*3. Obtain I_τ by taking the intersection of I_{re} and I_{fin}, then I_{fin} minus I_τ, and obtain connected regions I_{in} which can fill I_{re}.
*Step*4. In the model of the eight neighborhoods, in this study, each connected domain of I_{in} have been used to recover vessels that were denoised. The number of the connected domains before and after the recovering operation have been compared. If reduced, it implies that there are misclassified vascular bifurcations and crossovers in this noise region, and the vascular vessels should be recovered. On the contrary, if it increases, it shows that the noise region is real, and it should not be recovered.
*Step*5. Remove the connected domain that is less than 20 pixels to obtain a better optimized result.

The segmentation results usually consist of some small isolated regions caused by noise, and these regions are sometimes wrongly detected as vessels (18, 20). We used a post-processing method based on connected domains and removed less than or equal to 20 unconnected pixels considered as a non-vessel or a part of the background noise. So the unconnected non-vessel pixel was eliminated, and the thin connected vessel was preserved. After these steps, by identifying and recovering the vascular bifurcations and breakpoints, the continuity and the accuracy of vascular vessels segmentation results were improved. The post-processing operation gives a final resultant binary image.

EXPERIMENTAL RESULTS AND ANALYSIS

In our experiments, the proposed method was evaluated on the publicly available DRIVE database and STARE database. It is worth mentioning that the scarcity of artificial calibration image data is a major obstacle for medical image processing research and development. So, even if the number of images in each database is not big enough, the two databases still occupy a very important position in the field of blood vessel segmentation in fundus images. Since these databases contain ground truth maps manually segmented by different professionals, they have gained popularity in the field of retinal vessel segmentation. Experiments on those two databases also facilitate the comparison of the methods proposed in this article with other methods. Three different measures are used to evaluate the performance. Our experiments were tested with MATLAB in the environment of 2.5 Ghz Intel i5-3210 M CPU and 4 GB memory.

In order to verify the versatility and feasibility of the proposed method, we added a set of experiments, using our method to segment the fundus image in the publicly available CHASEDB1 databases. We used the same three metrics as the aforementioned experiment to evaluate performance. In addition, in order to use a more meaningful measure for the evaluation of the quality of pixel-wise segmentation and to compare with the recent literature to show the effectiveness of our proposed method, the Matthews correlation coefficient (MCC) is introduced (6, 34), the supplementary experiment was tested with MATLAB in the environment of a 2.5 Ghz Intel i5-10300H CPU and 16 GB memory.

Database

The DRIVE database consists of 40 color images taken by a Canon CR5 3CCD camera with a 45○ FOV which are divided into a training set and a test set, each containing 20 images. The size of each image is 768 × 584 pixels, with 8 bits per color channel, and the FOV is circular with 450 pixels in diameter. In the DRIVE database, there is a corresponding mask that delineates the FOV area and the binary vessel segmentation. The images in the test set have been segmented by two human observers, while the images in the training set have been segmented by one observer.

The STARE database consists of 20 color images, and half of the STARE database images contains signs of pathologies. A Topcon TRC-50 fundus camera at 35○ FOV was used to acquire the images. The size of each image is 700 × 605 pixels with 8 bits per color channel, and the FOV in the images is around 650 × 550 pixels. The images in the STARE database have been manually segmented by two different observers.

The CHASEDB1 database consists of 28 color images taken from the eyes of 14 school children. Usually, the first 20 images are used for training, and the remaining eight images are used for testing. The size of each image is 999 × 960 pixels, and the binary field-of-view (FOV) mask and segmentation ground truth are obtained by manual methods.

For the comparability of experimental data, the performance of the proposed method is measured on the test set of the DRIVE database and on the all images of the STARE database and CHASEDB1 database by comparing the automatically generated binary images with the ones that are segmented by the first observer as gold standard.

Evaluation Method

Each resulting binary image was compared with the corresponding gold standard by computing the four performance measurements: true positive (TP) is the number of pixels correctly classified as vessels, false positive (FP) is the number of pixels misclassified as vessels, true negative (TN) is the number of pixels correctly classified as backgrounds, and false negative (FN) is the number of pixels misclassified as backgrounds.

In order to evaluate the performance of our method and compare with state-of-the-art methods, we computed the measures known as accuracy (Acc), sensitivity (Se), and specificity (Sp) in **Table 2**.

The Acc of one image is a fraction of pixels representing the ratio of the total number of correctly classified pixels to the total number of pixels in the image FOV. Se is determined

TABLE 2 | Performance measures of vessel segmentation.

Performance measures	Description
Sensitivity(Se)	TP/(TP+FN)
Specificity(Sp)	TN/(TN+FP)
Accuracy(Acc)	(TP+TN)/(TP+FP+TN+FN)

TABLE 3 | Performance of proposed method (DRIVE).

	Time Consuming(s)	Se	Sp	Acc
Average	9.1852	0.7339	0.9847	0.9604
Maximum	10.1094	0.8455	0.9965	0.9696
Minimum	8.2813	0.5006	0.9737	0.9543

TABLE 4 | Performance of proposed method (STARE).

	Time consuming(s)	Se	Sp	Acc
Average	12.2773	0.8003	0.9705	0.9558
Maximum	15.5938	0.9250	0.9900	0.9764
Minimum	10.7118	0.6426	0.9313	0.9245

by dividing the number of pixels correctly classified as vessel pixels by the total vessel pixels in the manual segmentation; thus, Se denotes the ability of correctly identifying the vessel pixels. Sp is determined by dividing the number of pixels correctly classified as background pixels by the total background pixels in the manual segmentation; thus, Sp reflects the ability to detect non-vessel pixels.

In addition, we evaluated MCC indicators referring to Ricci et al. (41). The MCC (7, 21) is a more appropriate indicator of the accuracy of binary categorization in the case of unbalanced structures. The MCC counting transformation is defined as follows:

$$\mathrm{MCC} = \frac{(\mathrm{TP}/\mathrm{N} - \mathrm{S} \times \mathrm{P})}{\sqrt{\mathrm{P} \times \mathrm{S} \times (1 - \mathrm{S}) \times (1 - \mathrm{p})}} \tag{9}$$

where $\mathrm{N} = \mathrm{TN} + \mathrm{TP} + \mathrm{FN} + \mathrm{FP}$, $\mathrm{S} = (\mathrm{TP} + \mathrm{FN})/\mathrm{N}$ and $\mathrm{P} = (\mathrm{TP} + \mathrm{FP})/\mathrm{N}$.

Performance of the Proposed Method

Because the proposed method does not need to train a classifier, images in the training set of the DRIVE database have not been used. We tested on 20 fundus images from the test set of the DRIVE database. The total segmenting time is about 185 s, while each image takes about 9.19 s. The performance results of retinal vessel segmentation on the DRIVE database are shown in **Table 3**. The average *Acc*, *Se*, and *Sp* of proposed method are 0.9604, 0.7339, and 0.9847, respectively.

For the STARE database, we tested all 20 fundus images since the proposed method does not have to select images to train a classifier. The total segmenting time is about 242 s, while each image takes about 12.28 s. The performance results of retinal vessel segmentation on the STARE database are shown in **Table 4**. The average *Acc*, *Se*, and *Sp* of the proposed method are 0.9558, 0.8003, and 0.9705, respectively.

For the CHASEDB1 database, we tested all 28 fundus images for the same reason that the proposed method does not have to select images to train a classifier. The total segmenting time is about 251 s, while each image takes about 8.96 s. The average Acc, Se, and Sp of the proposed method are 0.9606, 0.6921, and 0.9842, respectively. We still introduced the Matthews correlation coefficient with a score of 0.5886. The performance results of retinal vessels segmentation on the CHASEDB1 database are shown in **Table 5**.

TABLE 5 | Performance of proposed method (CHASEDB1).

	Time consuming(s)	Se	Sp	Acc	MCC
Average	9.3600	0.5921	0.9842	0.9606	0.5886
Maximum	10.4271	0.7864	0.9959	0.9717	0.75710
Minimum	7.7188	0.4959	0.9313	0.9554	0.3970

The proposed method can generate the segmentation map corresponding to the fundus map in a short time, as shown in **Tables 3**–**5**. The processing time for segmenting DRIVE, STARE, and CHASEDB1 fundus map data using our method can reach 8.2318, 10.7188, and 7.7188s, respectively. This shows that the proposed method has real-time feasibility. In addition, our experiments were tested using MATLAB in different environments, and the test results are shown in **Tables 3**–**5**, which shows that our experiments are experimentally feasible.

Step-by-step representation of the proposed framework applied to images of DRIVE, STARE, and CHASEDB1 databases has been depicted in **Figures 3**–**5**, respectively. In the final segmentation result graph, looking closely at the segmentation of the vessel ends, it is obvious that the method used in this study has good effectiveness in distinguishing between vessel-like noise and thin vessels.

Comparative Experiment

We compared the proposed B-COSIRE filter approach with other state-of-the-art methods including supervised methods (the top six methods), unsupervised methods (in the middle), and five methods of deep learning (the last four methods) on DRIVE and STARE databases, respectively. The performance in the term of accuracy, sensitivity, and specificity is tabulated in **Tables 6**, **7**, respectively.

The proposed framework shows the almost highest results on the DRIVE images for supervised, unsupervised, and deep learning methods, with Acc = 0.9604, Se = 0.7339, and Sp = 0.9847. Our proposed technique also showed high efficiency in terms of sensitivity and specificity among all kinds of techniques on the STARE dataset. The accuracy Acc = 0.9558 also shows a great performance among the methods compared.

In the test experiments on DRIVE and STARE datasets, the proposed method consumed much lower time overhead than supervised methods and deep learning methods and also has advantages over some unsupervised methods. In the CHASEDB1 dataset supplementary experiment, we tested 28 fundus images; the total segmentation time was 251 seconds and the average time was 8.96 s, which shows that our time overhead is relatively low. Furthermore, in the experimental phase, our method does not need to use a large number of image annotations for training. This study used three lightweight public datasets but obtained relatively high test performance, as shown in **Tables 6**, **7**. It can

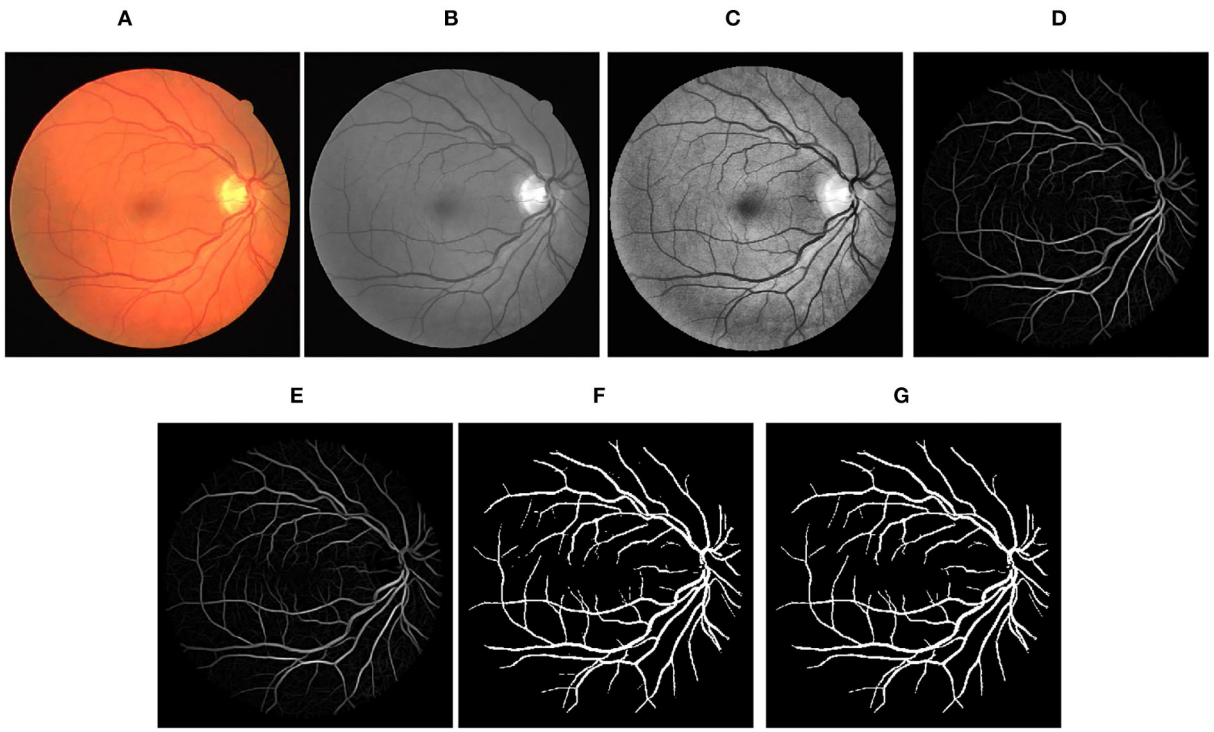


FIGURE 3 | Stepwise illustration of proposed method. **(A)** Color retinal fundus image from DRIVE database. **(B)** Green channel retinal fundus image. **(C)** Processing diagram after CLAHE. **(D)** Processing diagram after B-COSFIRE filters. **(E)** Processing diagram after morphological filters. **(F)** Processing diagram after binary threshold. **(G)** Final segmented image.

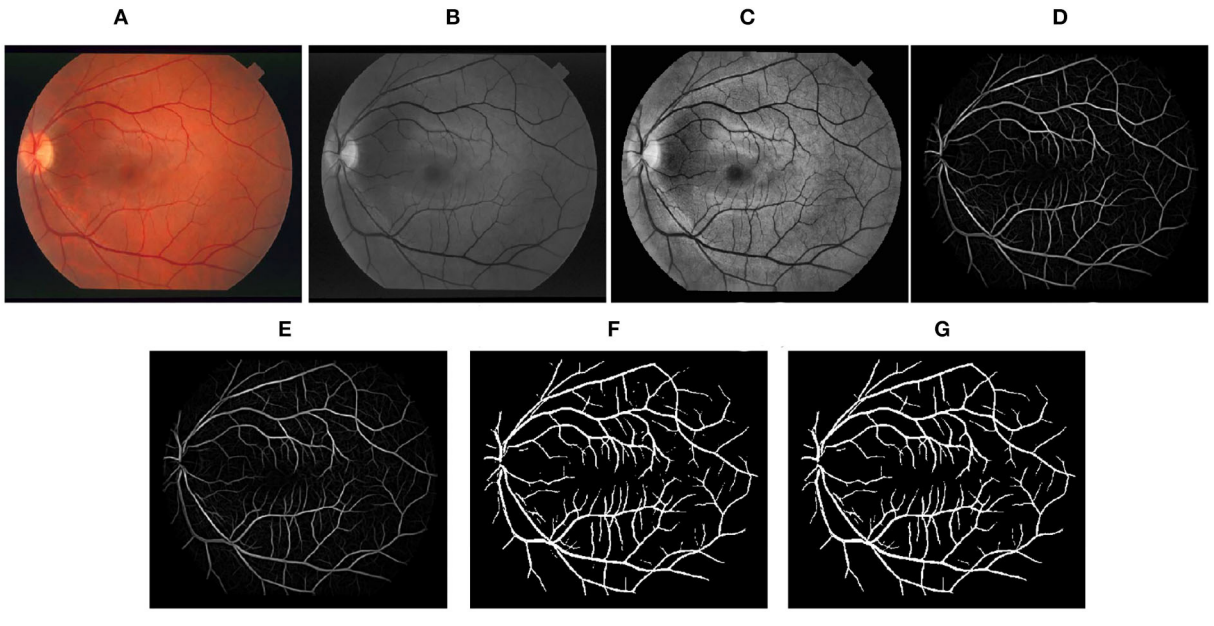


FIGURE 4 | Stepwise illustration of the proposed method. **(A)** Color retinal fundus image from STARE database. **(B)** Green channel retinal fundus image. **(C)** Processing diagram after CLAHE. **(D)** Processing diagram after B-COSFIRE filters. **(E)** Processing diagram after morphological filters. **(F)** Processing diagram after binary threshold. **(G)** Final segmented image.

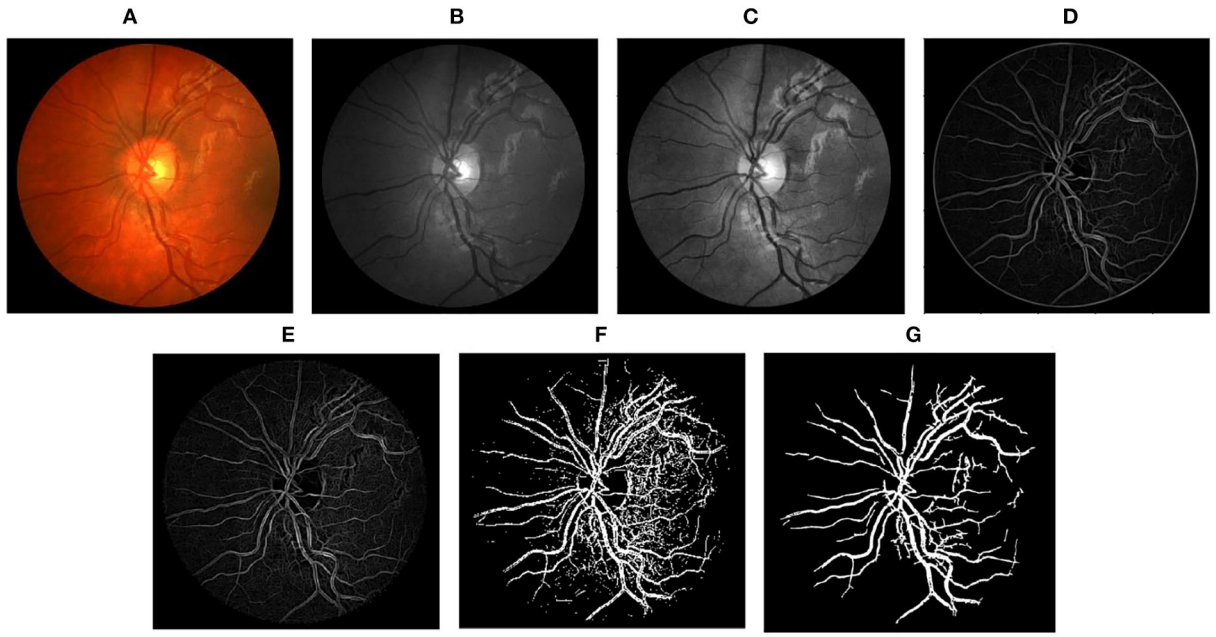


FIGURE 5 | Stepwise illustration of the proposed method. **(A)** Color retinal fundus image from CHASEDB1 database. **(B)** Green channel retinal fundus image. **(C)** Processing diagram after CLAHE. **(D)** Processing diagram after B-COSFIRE filters. **(E)** Processing diagram after morphological filters. **(F)** Processing diagram after binary threshold. **(G)** Final segmented image.

TABLE 6 | Comparison with other methods (DRIVE).

Method category	Reference	Year	Segmentation time	Se	Sp	Acc
Supervised	Soares et al. (10)	2006	3 min	0.7332	0.9782	0.9614
Supervised	Marin et al. (11)	2011	1.5 min	0.7067	0.9801	0.9588
Supervised	Fraz et al. (12)	2012	2 min	0.7406	0.9807	0.9747
Supervised	Aslani and Sarnel (13)	2016	-	0.7545	0.9801	0.9513
Supervised	Strisciuglio et al. (14)	2016	1.5 min	0.7731	0.9708	0.9453
Supervised	Zhu et al. (15)	2016	51 s	0.7462	0.9838	0.9618
Unsupervised	Al-Diri et al. (16)	2009	11 min	0.7282	0.9551	-
Unsupervised	Liu and Sun (17)	2010	13 min	-	-	0.9472
Unsupervised	Azzopardi et al. (3)	2015	10 s	0.7655	0.9704	0.9442
Unsupervised	Khan et al. (18)	2016	10.6 s	0.7155	0.9805	0.9579
Unsupervised	Bahadarkhan et al. (19)	2016	1.56 s	0.776	0.972	0.947
Unsupervised	Khan et al. (20)	2016	-	0.780	0.972	0.952
Unsupervised	Khan et al. (43)	2016	6.1 s	0.747	0.980	0.960
Unsupervised	Khan et al. (21)	2020	5.5 s	0.766	0.972	0.954
Deep learning	Melinsca et al. (25)	2015	-	0.7276	-	0.9466
Deep learning	Khalaf et al. (26)	2016	-	0.8467	0.9494	0.9403
Deep learning	Fu et al. (27)	2016	-	0.7294	-	0.9470
Deep learning	Cheng et al. (29)	2020	-	0.7676	0.9834	0.9559
Unsupervised	Proposed Method	2022	9.19 s	0.7339	0.9847	0.9604

be seen that the proposed method has extremely low overhead on the dataset.

As illustrated before, the proposed unsupervised method is better than other unsupervised methods. The performance of the proposed method is almost the same as that of the supervised methods, although is slightly worse than that of the best performance of deep learning methods. It is important that the proposed method is very efficient, while supervised methods and deep learning methods both consume much time in the process of classifier training. It proves the favorable applicability

TABLE 7 | Comparison with other methods (STARE).

Method category	Reference	Year	Segmentation time	Se	Sp	Acc
Supervised	Soares et al. (10)	2006	3 min	0.7207	0.9747	0.9480
Supervised	Marin et al. (11)	2011	1.5 min	0.6944	0.9819	0.9526
Supervised	Fraz et al. (12)	2012	2 min	0.7548	0.9763	0.9534
Supervised	Aslani et al. (13)	2016	-	0.7556	0.9837	0.9789
Supervised	Strisciuglio et al. (14)	2016	2.5 min	0.7668	0.9711	0.9545
Supervised	Zhu et al. (15)	2016	51 s	-	-	-
Unsupervised	Al-Diri et al. (16)	2009	11 min	0.7251	0.9681	-
Unsupervised	Liu and Sun (17)	2010	13 min	-	-	0.9567
Unsupervised	Azzopardi et al. (3)	2015	10 s	0.7716	0.9701	0.9497
Unsupervised	Khan et al. (18)	2016	10.6 s	0.7728	0.9649	0.9518
Unsupervised	Bahadarkhan et al. (19)	2016	1.56 s	0.895	0.939	0.935
Unsupervised	Khan et al. (20)	2016	-	0.745	0.74	0.957
Unsupervised	Khan et al. (43)	2016	6.1 s	0.778	0.966	0.951
Unsupervised	Khan et al. (21)	2020	5.5 s	0.792	0.997	0.996
Deep learning	Melinsca et al. (25)	2015	-	0.7276	-	0.9466
Deep learning	Khalaf et al. (26)	2016	-	0.8467	0.9494	0.9403
Deep learning	Fu et al. (27)	2016	-	0.7294	-	0.9470
Deep learning	Cheng et al. (29)	2020	-	-	-	-
Unsupervised	Proposed method	2022	12.28 s	0.8003	0.9705	0.9558

of the proposed method, which is fast and effective in the field of computer-aided diagnosis.

FINDINGS

In our work, we proposed a vessel segmentation method of fundus images. In the course of the experiment, we discovered some other issues that may be worth studying. First, in the image preprocessing process, the application of contrast enhancement algorithms to the image will not necessarily lead to better segmentation because some contrast enhancement algorithms will also enhance the background features (44), causing the segmented image to contain more noise. In supplementary experiment, we replaced the CLAHE algorithm in our proposed method with the GLM algorithm referring to Khan et al. (43) and found that the segmentation effect is comparable. We assume this is affected by the environment and data. This is also the direction of our next work. We look forward to making improvements in the method proposed by Khan et al. (43) to further optimize the segmentation method. Second, we found that there are white bars of noise in the blood vessels in part of the binarized images after segmentation. We were able to eliminate it through morphological operations, but the effect still has room for improvement. We believe that we can consider the method of multi-scale input fusion in future.

CONCLUSION

This article presents an improved unsupervised method for vascular segmentation on retinal color fundus images. The proposed method is based on the B-COSFIRE filter, through a series of operations including CLAHE, B-COSFIRE filters, morphological filters, and post-process to obtain final binary vessel images. The method is tested on the public databases, DRIVE, STARE, and CHASEDB1. The proposed method requires little processing time (around 9 s for each image in DRIVE and CHASEDB1, 12 s for each image in STARE) and results in the average accuracy, sensitivity, and specificity of 0.9604, 0.7339, and 0.9847 for DRIVE database; 0.9558, 0.8003, and 0.9705 for STARE database; and 0.6921, 0.9842, and 0.9606 for CHASEDB1 database, respectively. In general, the method used in this study has the following advantages: 1. low time overhead and low dataset overhead, 2. good versatility in the field of computer-aided diagnosis, and 3. a relatively high segmentation effect while maintaining a relatively low time overhead. Through the analysis of the experimental results, it is proved that the method proposed in this article is cutting-edge and effective in the field of retinal blood vessel segmentation. In conclusion, the proposed method can be employed for computer-aided diagnosis, disease screening, and any other circumstances that require fast delineation of blood vessels. It may help us prevent many related diseases, such as diabetes and glaucoma. The direction of our future investigation should be done by configuring specific COSFIRE filters of different shapes with other methods of image processing to solve lesion extraction in retinal fundus images.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://blogs.kingston.ac.uk/retinal/

chasedb1/; http://www.isi.uu.nl/Research/Databases/DRIVE/download.php.

AUTHOR CONTRIBUTIONS

WL, YX, and HH are the experimental designers and executors of this study, completed the data analysis, and wrote the first draft of the paper. CZ, HW, and ZL participated in the experimental design and analysis of the experimental results. AS is the project designer and director, who directed the experimental design, data analysis, thesis writing, and revision. All authors read and agree to the final text.

FUNDING

This work is supported by the Scientific and Technological Innovation Leading Plan of High-tech Industry of Hunan Province (2020GK2021), the National Natural Science Foundation of China (61702559), the Research on the Application of Multi-modal Artificial Intelligence in Diagnosis and Treatment of Type 2 Diabetes under Grant No. 2020SK50910, the International Science and Technology Innovation Joint Base of Machine Vision and Medical Image Processing in Hunan Province (2021CB1013), and the Natural Science Foundation of Hunan Province (No. 2022JJ30762).

REFERENCES

1. Yusup M, Chen XY. Epidemiology survey of visual loss. *Int J Ophthalmol.* (2010) 10:304–7. doi: 10.3969/j.issn.1672-5123.2010.02.033
2. Resnikoff S, Pascolini D, Etya'Ale D, Kocur I, Pararajasegaram R, Pokharel G P, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ*. (2004) 82:844–51. doi: 10.0000/PMID15640920
3. Azzopardi G, Strisciuglio N, Vento M, Petkov N. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Med Image Anal.* (2015) 19:46–57. doi: 10.1016/j.media.2014.08.002
4. Azzopardi G, Petkov N. A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model. *Biol Cybern.* (2012) 106:177–89. doi: 10.1007/s00422-012-0486-6
5. Azzopardi G, Azzopardi N. Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Trans Pattern Anal Mach Intell.* (2012) 35:490–503. doi: 10.1109/TPAMI.2012.106
6. Khan KB, Khaliq AA, Jalil A, Iftikhar MA, Ullah N, Aziz MW, et al. A review of retinal blood vessels extraction techniques: challenges, taxonomy, and future trends. *Pattern Anal Appl.* (2018) 22:767–802. doi: 10.1007/s10044-018-0754-8
7. Owen CG, Rudnicka AR, Mullen R, Barman SA, Monekosso D, Whincup PH, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the Computer-Assisted Image Analysis of the Retina (CAIAR) program. *Invest Ophthalmol Vis Sci.* (2009) 50:2004–10. doi: 10.1167/iovs.08-3018
8. Mittal M, Iwendi C, Khan S, Javed AR. Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using Levenberg-Marquardt neural network and gated recurrent unit for intrusion detection system. *Trans Emerg Telecommun Technol*. (2021) 32: e3997. doi: 10.1002/ett.3997
9. Strisciuglio N, Azzopardi G, Vento M, Petkov N. *Multiscale blood vessel delineation using B- COSFIRE filters. International Conference on Computer Analysis of Images and Patterns*. Valletta: Springer, Cham. (2015).
10. Soares J, Leandro J, Cesar RM, Jelinek HF, Cree M. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging.* (2006) 25:1214–22. doi: 10.1109/TMI.2006.879967
11. Marin D, Aquino A, Gegundez-Arias ME, Bravo JM. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imaging.* (2011) 30:146. doi: 10.1109/TMI.2010.2064333
12. Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, et al. Blood vessel segmentation methodologies in retinal images – A survey. *Comput Methods Prog Biomedi.* (2012) 108:407–33. doi: 10.1016/j.cmpb.2012.03.009
13. Aslani S, Sarnel H. A new supervised retinal vessel segmentation method based on robust hybrid features. *Biomed Signal Process Control.* (2016) 30:1–12. doi: 10.1016/j.bspc.2016.05.006
14. Strisciuglio N, Azzopardi G, Vento M, Petkov N. Supervised vessel delineation in retinal fundus images with the automatic selection of B-COSFIRE filters. *Mach Vision Appl.* (2016) 27:1137–49. doi: 10.1007/s00138-016-0781-7
15. Zhu C, Zou B, Xiang Y, Cui J, Hui WU. An ensemble retinal vessel segmentation based on supervised learning in fundus images. *Chin J Electron.* (2016) 25:503–11. doi: 10.1049/cje.2016.05.016
16. Al-Diri B, Hunter A, Steel D. An Active Contour Model for Segmenting and Measuring Retinal Vessels. *IEEE Trans Med Imaging*. (2009) 28:1488–97. doi: 10.1109/TMI.2009.2017941
17. Liu I, Sun Y. Recursive tracking of vascular networks in angiograms based on the detection- deletion scheme. *IEEE Trans Med Imaging.* (1993) 12:334–41. doi: 10.1109/42.232264
18. Khan KB, Khaliq AA, Shahid M, B-COSFIRE filter and VLM based retinal blood vessels segmentation and denoising. In: *International Conference on Computing*. Quetta: IEEE. (2016).
19. Bahadarkhan K, Khaliq AA, Shahid M. A morphological hessian based approach for retinal blood vessels segmentation and denoising using region based otsu thresholding. *PLoS ONE.* (2016) 11:1646–2744. doi: 10.1371/journal.pone.0158996
20. Khan KB, Khaliq AA, Shahid M, Khan S. An efficient technique for retinal vessel segmentation and denoising using modified isodata and Clahe. *Int Islamic Univ Malays Eng J.* (2016) 17:31–46(16). doi: 10.31436/iiumej.v17i2.611
21. Khan KB, Siddique MS, Ahmad M, Mazzara M, Khattak DK. A hybrid unsupervised approach for retinal vessel segmentation. *BioMed Res Int*. (2020) 2020. doi: 10.1155/2020/8365783
22. Chen C, Chuah JH, Ali R, Wang Y. Retinal vessel segmentation using deep learning: a review. *IEEE Access*. (2021) 9:111985–2004. doi: 10.1109/ACCESS.2021.3102176
23. Dhiman P, Kukreja V, Manoharan P, Kaur A, Kamruzzaman MM, Dhaou IB, et al. A novel deep learning model for detection of severity level of the disease in citrus fruits. *Electronics.* (2022) 11:495. doi: 10.3390/electronics11030495
24. Gadamsetty S, Ch R, Ch A, Iwendi C, Gadekallu TR. Hash-based deep learning approach for remote sensing satellite imagery detection. *Water.* (2022) 14:707. doi: 10.3390/w14050707
25. Melinsca M, Prentasic P, Loncaric S. Retinal vessel segmentation using deep neural networks. In: *International Conference on Computer Vision Theory and Applications*. Coimbatore. (2015).
26. Khalaf AF, Yassine IA, Fahmy AS. Convolutional neural networks for deep feature learning in retinal vessel segmentation. IEEE International Conference on Image Processing. *IEEE.* (2016) 10:385–8. doi: 10.1109/ICIP.2016.7532384
27. Fu H, Xu Y, Wong D, Jiang L. Retinal vessel segmentation via deep learning network and fully- connected conditional random fields. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI 2016)*. Prague: IEEE. (2016).
28. Yue K, Zou B, Chen Z, Liu Q. Retinal vessel segmentation using dense U-net with multiscale inputs. *J Med Imaging.* (2019) 6:1. doi: 10.1117/1.JMI.6.3.034004
29. Cheng Y, Ma M, Zhang L, Jin C J, Zhou Y. Retinal blood vessel segmentation based on Densely Connected U-Net. *Math Biosci Eng.* (2020) 17:3088–108. doi: 10.3934/mbe.2020175

30. Ooi AZH, Embong Z, Abd Hamid AI, Zainon R, Wang SL, Ng TF, et al. Interactive blood vessel segmentation from retinal fundus image based on canny edge detector. *Sensors.* (2021) 21:1424–8220. doi: 10.3390/s21196380
31. Li Z, Jia M, Yang X, Xu M. Blood vessel segmentation of retinal image based on dense-U-net network. *Micromachines (Basel).* (2021) 12:1478. doi: 10.3390/mi12121478
32. Zhai ZL, Feng S, Yao LY, Li PH. Retinal vessel image segmentation algorithm based on encoder- decoder structure. *Multimedia Tools Appl.* (2022) 13167:157. doi: 10.1007/s11042-022-13176-5
33. Azzopardi G, Petkov N. A shape descriptor based on trainable COSFIRE filters for the recognition of handwritten digits. In: *International Conference on Computer Analysis of Images & Patterns*. Berlin, Heidelberg: Springer. (2013).
34. Azzopardi G, Petkov N. Automatic detection of vascular bifurcations in segmented retinal images using trainable COSFIRE filters. *Pattern Recognit Lett*. (2013) 34:922–33. doi: 10.1016/j.patrec.2012.11.002
35. Azzopardi G, Greco A, Vento M. Gender recognition from face images with trainable COSFIRE filters. In: *IEEE International Conference on Advanced Video & Signal Based Surveillance. IEEE.* Colorado Springs. (2016).
36. Gecer B, Azzopardi G, Petkov N. Color-blob-based COSFIRE filters for object recognition. *Image Vision Comput*. (2017) 57:165–74. doi: 10.1016/j.imavis.2016.10.006
37. Guo J, Shi C, Azzopardi G, Petkov N. Recognition of Architectural and Electrical Symbols by COSFIRE Filters with Inhibition. In: *Computer analysis of Images and Patterns*. Valletta: Springer, Cham. (2015).
38. Guo J, Shi C, Azzopardi G, Petkov N. Inhibition-augmented trainable COSFIRE filters for keypoint detection and object recognition. *Mach Vision Appl.* (2016) 27:1197–211. doi: 10.1007/s00138-016-0777-3
39. Niemeijer M, Staal JJ, Ginneken BV, Loog M, Abramoff MD. Comparative study of retinal vessel segmentation methods on a new publicly available database. *Proc SPIE.* (2004) 5370:648–56. doi: 10.1117/12.535349
40. Mendonca AM, Campilho A. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging*. (2006) 25:1200–13. doi: 10.1109/TMI.2006.879955
41. Ricci E, Perfetti R. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging.* (2007) 26:1357. doi: 10.1109/TMI.2007.898551
42. Fadzil M, Nugroho HA, Nugroho HA, znita IL. Contrast Enhancement of Retinal Vasculature in Digital Fundus Image. In: *International Conference on Digital Image Processing IEEE.* Bangkok. (2009).
43. Khan KB, Khaliq AA, Shahid M. A novel fast GLM approach for retinal vascular segmentation and denoising. *J Inf Sci Eng.* (2017) 33:1611–27. doi: 10.6688/JISE.2017.33.6.14
44. Imran A, Li J, Pei Y, Yang JJ, Wang Q. Comparative analysis of vessel segmentation techniques in retinal images. *IEEE Access.* (2019) 7:114862–87. doi: 10.1109/ACCESS.2019.2935912

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Xiao, Hu, Zhu, Wang, Liu and Sangaiah. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 27 September 2022
DOI 10.3389/fpubh.2022.947223

Check for updates

OPEN ACCESS

EDITED BY
Ali Kashif Bashir,
Manchester Metropolitan University,
United Kingdom

REVIEWED BY
Yalin Wu,
Lushan Botanical Garden (CAS), China
Kechao Wang,
Harbin University, China
Ping Zhang,
Lanzhou University of
Technology, China

*CORRESPONDENCE
Bin Zhang
zhangbinzjjy@163.com

SPECIALTY SECTION
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

RECEIVED 18 May 2022
ACCEPTED 17 August 2022
PUBLISHED 27 September 2022

CITATION
Chen D and Zhang B (2022) A home fitness satisfaction model for Chinese residents during the COVID-19 pandemic based on SEM analysis.
Front. Public Health 10:947223.
doi: 10.3389/fpubh.2022.947223

COPYRIGHT
© 2022 Chen and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A home fitness satisfaction model for Chinese residents during the COVID-19 pandemic based on SEM analysis

Dexu Chen[1] and Bin Zhang[2]*

[1]School of Physical Education, Shandong University, Jinan, China, [2]Zhejiang College of Construction, Hangzhou, China

In response to the COVID-19 epidemic, China proposes to take this epidemic as an opportunity to further cultivate residents' healthy living habits and meet their demand for healthy living. Based on relevant research reviews and theoretical analysis, the concept model of home fitness satisfaction for Chinese residents is constructed. The research hypothesis is tested by SEM, and the conceptual model is modified. Finally, the model of home fitness satisfaction for Chinese residents is obtained. The results show that during the epidemic period, Chinese residents' enthusiasm for home fitness remains unchanged, online services provide powerful social support, and the home fitness environment still has an enhanced space.

KEYWORDS

home fitness, satisfaction, healthcare, structural equation, COVID-19

Problem statement

The COVID-19 epidemic broke out all over the world, which imposed a near-total lockdown on citizens in China. In this war without smoke, staying at home played an important part. During the epidemic period, the General Office of the State General Administration of Sport issued the notice on vigorously promoting scientific home-based fitness methods, requiring local sports departments to actively promote scientific home-based fitness and promote people's health with the cooperation of local party committees and governments. Media platforms at all levels and of all kinds have launched home-based fitness programs, and home-based fitness has become a hot topic. The number of search results for "epidemic home-based fitness" and "epidemic home-based exercise" on Baidu was 18,300,000 and 43,900,000, respectively. On the one hand, the outbreak of the epidemic caused great mental trauma for people, forcing people to re-examine the importance of a healthy lifestyle, which helps people make greater progress between the residents of sports and healthy cognition. Meanwhile, the value function of sports fitness to improve people's physical fitness and promote physical and mental health is manifested again in this special period. On the other hand, affected by the epidemic, sports departments at all levels, schools, as well as sports and fitness institutions have launched online fitness courses and sports and fitness live broadcasts. Online fitness clocking and other services have attracted high attention. However, the flash-in-the-pan craze of home fitness is not what society needs. How to continue to promote home

fitness services, how to make more people more convenient to participate in in-home fitness, and how to make use of Internet technology to promote the reform of social sports services are the key points of our further study. Therefore, it is necessary to study home fitness from the perspective of residents' satisfaction to enhance the effect of home fitness, strengthen the weak links of home fitness, and promote the health of the people. Based on the investigation of residents' satisfaction with home fitness and its influencing factors during the epidemic period, this study builds a model of residents' satisfaction and analyzes the influencing factors of residents' home fitness during the epidemic period to improve the effect of home fitness and promote national fitness.

Infection with novel coronavirus pneumonia does not necessarily have a physical impact, as in a latent phase, but if it is in the onset phase, the condition is of varying severity, and the physical impact is different and cannot be generalized. (1) Latency. If the following infection with novel coronavirus pneumonia is found at a latent stage of the condition, usually without obvious symptoms, it may be detected at the time nucleic acid testing is performed, at which point there will be no apparent effect on the body. (2) Onset stage. Less severe: if the new coronavirus pneumonia develops after the onset, the condition is relatively mild and may present with fever, dry cough, and other symptoms, but also with generalized weakness, some people will have symptoms such as runny nose and sneezing, should actively visit the hospital for supportive care, make appropriate exercise to improve their physical fitness, the condition generally recovers slowly and does not cause many effects on their physical health; Severe disease: if the new coronavirus pneumonia continues to progress, is not treated promptly, causes more damage to the body, may present with symptoms such as dysgeusia, anosmia, or even dyspnea, and some may present with hypoxemia, sepsis, shock, or acute respiratory distress syndrome, as well as multiple organ failure.

Model construction and research hypothesis

The study of "satisfaction" is coined in the study of "customer satisfaction" in economic problems. It is generally believed in the academic area that customer satisfaction is the psychological reaction of customers after their needs are met, and it is the evaluation of customers of the product or the degree to which it meets their own needs (1). In the 1990's, the American Customer Satisfaction Model (ACSI) and the European Customer Satisfaction Model (ECSI) were widely adopted, and in the 21st century, the Chinese Customer Satisfaction Model (CCSI) was created (2). The satisfaction model can reveal the relationship between various factors affecting satisfaction, and a structural equation is often used to verify and adjust the satisfaction model (3). Structural equations provide a rigorous scientific procedure for quantitative research of abstract concepts and enable researchers to test research hypotheses through scientific statistical analysis. In this study, structural equations were used to verify and build a satisfaction model, and the residents' satisfaction with home fitness and its influencing factors were included in the same model, and the relationship among multiple factors was processed and analyzed at the same time. Compared with the simple descriptive analysis, the structural equation model can better explain the residents' satisfaction with home fitness and its influencing factors during the epidemic period.

Conceptual model construction

Through comparative analysis, the American customer satisfaction model and China customer satisfaction index (ACSI) model (CSCI) in the "customer perceived value (4), customer expectations, and quality awareness," are closer to the residents' home fitness satisfaction variables such as model building, so this study is based on ACSI and CSCI built during the outbreak of residents' fitness satisfaction model at home. In the model, the intention to exercise at home, the perception of social support, the perception of the environment of exercise at home, the residents' physical literacy, and the gain of exercise at home are the cause variables, while the residents' satisfaction with exercise at home is the target variable. Figure 1 is the conceptual model of residents' satisfaction with home fitness during the epidemic period.

The concept of home fitness intention comes from the survey of residents' sports awareness and draws on the concept of "customer expectation," which is the expectation of customers for the products or services that will happen. The intention of exercising at home during the epidemic period refers to the residents' expectation to participate in the exercise at home during the epidemic period, including the idea of exercising at home, the desire to exercise with others, the expectation to bring pleasure to the exercise at home, and the expectation to release pressure through the exercise at home (5). Previous studies have shown that residents' intention to exercise at home has a positive impact on satisfaction with exercise at home, perception of social support for sports, and perception of support for sports environment. This study assumes that residents' intention to exercise at home has a positive impact on their perception of social support and fitness environment at home, residents' physical literacy, and residents' satisfaction with sports (6).

The perception of social support refers to the quality perception in the customer satisfaction model, which refers to the subjective feelings of customers when they receive products or services. In this study, the perception of social support refers to the support that residents received from the government, schools, media, and other platforms on home fitness during the period of the epidemic, as well as the online or offline

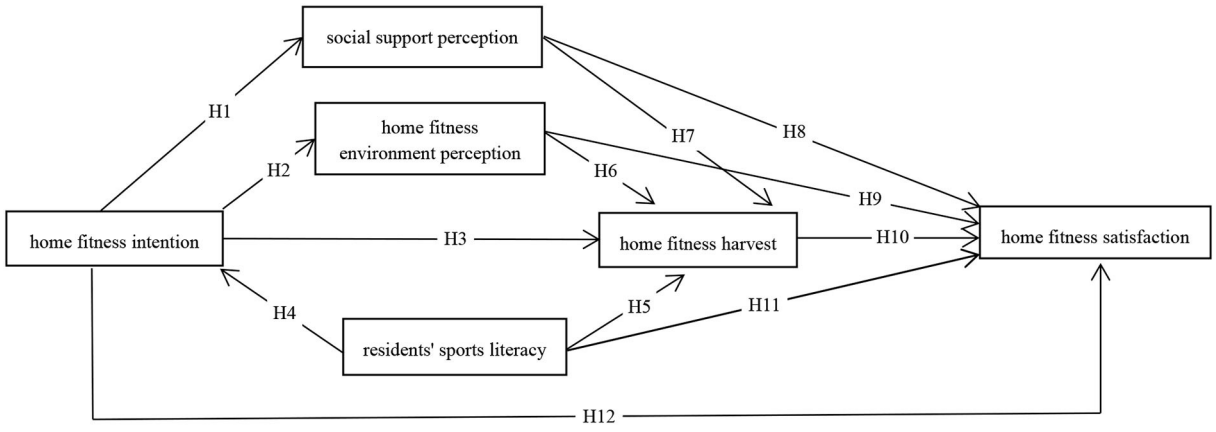


FIGURE 1
Conceptual model of residents' satisfaction with home fitness during the epidemic period.

home fitness guidance from others. This support can be either physical, such as mailing fitness books and equipment, or virtual, such as online guidance and a fitness clock. In this study, the perception of social support was mainly assessed by residents' evaluation of the above support methods and contents. In previous studies, researchers believe that quality perception directly affects customers' evaluation of products or services and is an important factor affecting customer satisfaction. Therefore, this study hypothesizes that the perception of social support has a positive impact on residents' sports harvest and satisfaction with sports (7).

The perception of a home fitness environment also comes from the concept of "quality perception." Home fitness environment perception refers to residents' evaluation of the size and comfort of the home fitness space, the type and quantity of equipment, and the home fitness atmosphere during their stay at home. The home-based fitness environment is the material guarantee for residents to participate in home-based fitness, which affects residents' home-based fitness to a great extent (8). This study hypothesizes that the perception of the home fitness environment plays a positive role in residents' home fitness harvest and satisfaction with home fitness.

Residents' physical literacy mainly refers to the acquired physical skills, physical knowledge, and physical behavior (9). To a certain extent, residents' physical literacy reflects their ability to participate in physical activities as well as their frequency of participation in physical activities. It is also an important indicator of whether residents can carry out independent scientific fitness tests. This study hypothesizes that residents' sports literacy has a positive effect on residents' sports harvest and satisfaction.

The home-based fitness harvest draws the concept of "customer value perception." Customer value perception refers to the overall evaluation of the product or service made by customers based on all their efforts and gains. The perception of home fitness gains refers to the evaluation of residents' physical and psychological feelings after home fitness practice. This feeling can be the physical relaxation of the exercise subject and psychological pleasure but can also be the emotional promotion between individuals participating in the exercise (10). The research believes that the gain of home fitness is an important factor in promoting residents' repeated participation in in-home fitness and that the gain of home fitness affects the satisfaction of home fitness to a large extent. The study hypothesizes that home-based fitness gains have a positive effect on home-based fitness satisfaction (11).

Residents' satisfaction with home fitness refers to residents' satisfaction with the "services" provided by society and their surroundings in the process of participating in in-home fitness. It is the psychological reaction of residents after they are satisfied with home fitness. Residents' satisfaction with home fitness is the target variable of this study. This study hypothesizes that residents' satisfaction with home fitness is affected by their intention to do home fitness, perception of social support, perception of the home fitness environment, residents' physical literacy, and gain from home fitness.

Research objectives and methodology

In this study, 1,052 residents in 30 provinces and cities were surveyed by a self-designed questionnaire, and the data were collected to verify the conceptual model. As the subjects involved in the questionnaire were somewhat affected by the outbreak, the questionnaire channels of this study could only communicate through online channels such as the Internet and telephone (subjects completed questionnaires at home).

TABLE 1 List of demographic characteristics of subjects ($n = 1,052$).

Variable	Frequency	Sample proportion (%)	Variable	Frequency	Sample proportion (%)
Gender			**Professional**		
Male	557	52.9	Cadres of state organs	95	9.0
Female	495	47.1	Employees of enterprises and public institutions	284	27.0
Age (years)			Professional and technical personnel	92	8.7
18–25	275	26.1	Production personnel in agriculture, forestry, and animal husbandry	126	12.0
25–39	266	25.3	Self-employed	74	7.0
40–49	234	22.2	Students	297	28.2
50–59	181	17.2	Other	84	8.0
60 and above	96	9.1	Regional		
Record of formal schooling	259	24.6	In the east	412	39.2
Senior high school and below					
The university of	617	58.7	In the middle	386	36.7
Postgraduate or above	176	16.7	In the west	254	24.1

This study refers to the statistical table of the 2015 National Population Sample Survey published by the National Bureau of Statistics.

TABLE 2 Consistency analysis of each dimension.

The dimension	Cronbach's α coefficient
Home fitness intention	0.916
Perception of social support	0.904
Home fitness environment perception	0.903
Residents' physical literacy	0.925
Home fitness gains	0.929
Satisfaction with home fitness	0.936

Research objectives

In this study, a total of 1,194 questionnaires were sent out, and 1,052 were effective, with a rate of 88.1%, including 557 males, accounting for 52.7%, and 495 females, accounting for 47.1%. A total of 275 people aged 18–25 years accounted for 26.1%; 266 people aged 26–39 years accounted for 25.3%; 234 people aged 40–49 years accounted for 22.2%; 181 people aged 50–59 years accounted for 17.2%; 96 people aged 60 years and above accounted for 9.1%. A total of 412 people in the eastern part accounted for 39.2%, 386 people in the central part accounted for 36.7%, and 254 people in the western part accounted for 24.1%. There are 259 people with a high school education or less, accounting for 24.6%; 617 people with a college education, accounting for 58.7%; 176 people with a master's degree or above, accounting for 16.7%. In total, 95, or 9%, were cadres of state organs; 284, or 27%, were employees of enterprises and public institutions; 92, or 8.7%, were professional technicians; 126, or 12%, were production workers in agriculture, forestry, and animal husbandry; 74, or 7%, were self-employed; 297, or 28.2%, were students; and 84, or 8%, were others. Table 1 is the list of demographic characteristics of subjects ($n = 1,052$).

Research methodology

The compilation of the questionnaire in this study is based on two aspects: first, by referring to the Chinese citizens' sports awareness survey report (2018) and related literature on the home fitness satisfaction survey, and combining it with the reality of home fitness during the epidemic, the interview outline was formulated [summing up the dimensions of the questionnaire of residents' satisfaction with home fitness during the period of the epidemic (participation degree of home fitness, the content of home fitness, space of home fitness venue, time, and frequency of participation in-home fitness)] and the second is to draw on the American Customer Satisfaction Model (ASCI) and the Chinese Customer Satisfaction Model (CSCI), formulate the contents of the questionnaire through interviews and revisions, and finally, determine the causal variables of the model based on home

TABLE 3 Correlation coefficients of factors in the questionnaire of satisfaction with home fitness and between factors and the total score.

Satisfaction with home fitness	Participation	Content	Floor space	Time and frequency	Total score
Participation	1	-	-	-	-
Content	0.708**	1	-	-	-
Floor space	0.714**	0.782**	1	-	-
Time and frequency	0.677**	0.826* *	0.742* *	1	-
Total score	0.756**	0.914**	0.905**	0.896**	1

**Is a statistical symbol.

TABLE 4 Correlation coefficients between factors and total scores of the questionnaire on influencing factors of residents' satisfaction with home fitness during the period of the epidemic.

	Home Fitness Intention	Social support	The sports environment	Sports accomplishment	Sports to harvest	Total score
Home fitness intention	1	-	-	-	-	-
Perception of social support	0.632**	1		-	-	-
Perception of sports environment	0.698**	0.612**	1	-	-	-
Residents' physical literacy	0.801**	0.624**	0.751**	1	-	-
Sports harvest for residents	0.813**	0.798**	0.794**	0.815**	1	-
Total score	0.886**	0.864**	0.881**	0.856**	0.909**	1

**Is a statistical symbol.

fitness intention, social sports culture support perception, home fitness environment perception, home fitness gain, and residents' sports literacy.

Based on the above content, writing the questionnaire on residents' fitness degree of satisfaction and its influencing factors at home during the outbreak of the COVID-19 epidemic, then eight experts reviewed the content of the questionnaire, the first draft of the questionnaire is identified after three rounds of modification, and try on a small scale questionnaire, with the respondents' feedback to the questionnaire of subject setting, the expression made revisions, the rationality and scientificity of the questionnaire are guaranteed. The revised questionnaire is divided into three parts. The first part is the demographic survey of the respondents, including gender, age, occupation, education background, and region. The second part is the questionnaire to assess residents' satisfaction with home fitness during the epidemic period. It uses 5-point Likert scale and sets 16 questions from four aspects, namely, sports participation, content, space, and time-frequency of home fitness. The third part is the questionnaire on influencing factors of residents' satisfaction with home fitness during the epidemic period. The 5-point Likert scale was also used to set 20 questions from five aspects, namely, home fitness intention, perception of social sports culture support, perception of home fitness environment, home fitness gain, and residents' physical literacy. SPSS22.0 and AMOS 24.0 were used for data processing.

Research results and analysis

Conceptual model analysis

Questionnaire reliability analysis mainly tests the consistency of each dimension of the questionnaire, which is generally measured by Cronbach's α coefficient. The consistency analysis of each dimension of this study is shown in Table 2. The Cronbach's α coefficient of each dimension is above 0.903, indicating that the questionnaire has high reliability.

Correlation analysis was used to verify the structural validity of the questionnaire. The correlation coefficient between the factors of residents' satisfaction with home fitness and the total score ranged from 0.756 to 0.914, showing a high correlation, and the correlation coefficient between the factors ranged from 0.677 to 0.826, showing a moderate correlation, as shown in Table 3.

The correlation coefficient between the influencing factors of home fitness satisfaction and the total score ranged from 0.856 to 0.909, showing a high correlation, and the correlation coefficient between each factor ranged from 0.612 to 0.813, showing a moderate correlation, as shown in Table 4. It shows that all factors are correlated with each other and also have some independence, which can reflect the contents of the questionnaire to be investigated. Based on the above analysis results, it can be concluded that the questionnaire on residents' satisfaction with home fitness and its influencing factors during

TABLE 5 The confirmatory factor analysis of residents' satisfaction with the home fitness questionnaire during the epidemic period showed that the fitting degree met the requirements.

Indicators	Squared/df	GFI	RMSEA	RMR	CFI	NFI	NNFI
The judgment standard	<3	>0.9	<0.10	<0.05	>0.9	>0.9	>0.9
Value	2.544	0.959	0.090	0.046	0.928	0.916	0.815

TABLE 6 Fit degree of influencing factors of residents' satisfaction with home fitness during the epidemic period.

Indicators	Squared/df	GFI	RMSEA	RMR	CFI	NFI	NNFI
The judgment standard	<3	>0.9	<0.10	<0.05	>0.9	>0.9	>0.9
Value	3.043	0.917	0.040	0.024	0.905	0.935	0.840

the epidemic period has good structural validity. Table 5 shows the confirmatory factor analysis of residents' satisfaction with the home fitness questionnaire during the epidemic period, showing that the fitting degree met the requirements.

The study conducted correlation analysis on the questionnaire of influencing factors of residents' satisfaction with home fitness during the period of the epidemic, and the fitting degree met the requirements. Table 6 shows the fit degree of influencing factors of residents' satisfaction with home fitness during the epidemic period.

The questionnaire score of each subject, whether in the influence factors, home fitness intention and social support perception problem is higher, shows that the participation of residents at the home had a higher fitness intention at home, and on the outbreak during the social from all walks of life support for home fitness is more satisfied, and home fitness literacy, environmental awareness, and home fitness score are not very high; it shows that there are still a large number of residents who do not have the habit of home exercise and enough fitness knowledge. They are not very satisfied with the overall environment of home exercise, and the gain of home exercise is not very great. Table 7 shows the average value, standard deviation, and average score for each dimension.

As can be seen from the satisfaction questionnaire, residents' sports participation scores are high, content and site equipment scores are average, and time and frequency scores are the lowest. It shows that the vast majority of residents participate in-home fitness but are not very satisfied with the content of exercise, exercise facilities, and equipment, and fail to form regular home fitness behavior.

TABLE 7 The average value, standard deviation, and the average score of each dimension.

The dimension	The average	The standard deviation	The topic split
Home fitness intention	14.88	2.41	3.72
Perception of social support	14.56	2.52	3.64
Home fitness environment perception	12.44	2.56	3.11
Residents' physical literacy	12.72	2.68	3.18
Home fitness gains	12.28	2.41	3.07
Participation	15.12	2.87	3.78
Content	12.56	2.65	3.14
Space equipment	12.16	2.51	3.04
Time and frequency	11.60	2.11	2.90

Hypothesis test

After a reliability and validity test and confirmatory factor analysis, the hypothesized model of this study was verified. After running the software, it is found that the fitting degree of the model does not meet the standard, and the path analysis results show that the hypotheses of H6 and H12 are not valid. Further modifications to the model are needed. Table 8 shows the test of significance of path coefficient of structural variables of a conceptual model.

Model modification

It is assumed that the reason why H6 is not completely true may be that some residents live in a poor home fitness environment, so residents do not feel the obvious impact of the home fitness environment on the harvest of home fitness. But given the outbreak period, residents' activity limitations, difficulty in obtaining normal home health supplies and equipment being unable to create a good fitness environment at home, and parts of the home fitness behavior, such as rope skipping, setting-up exercise, and treadmill running, is

TABLE 8 Test of significance of path coefficient of structural variables of a conceptual model.

Assuming that	Path to the relationship between			P	Path coefficient	Conclusion 1 $P < 0.05$	Conclusion 2 $P < 0.001$)
H1	Home fitness intention	->	Perception of social support	***	0.31	Support	Support
H2	Home fitness intention	->	Home fitness environment perception	***	0.32	Support	Support
H3	Home fitness intention	->	Residents harvest	***	0.37	Support	Support
H4	Residents' physical literacy	->	Home fitness intention	***	0.50	Support	Support
H5	Residents' physical literacy	->	Residents harvest	***	0.54	Support	Support
H6	Home fitness environment perception	->	Residents harvest	0.047	0.15	Support	Does not support
H7	Perception of social support	->	Residents harvest	***	0.36	Support	Support
H8	Perception of social support	->	Residents' satisfaction	***	0.44	Support	Support
H9	Home fitness environment perception	->	Residents' satisfaction	***	0.34	Support	Support
H10	Residents harvest	->	Residents' satisfaction	***	0.39	Support	Support
H11	Residents' physical literacy	->	Residents' satisfaction	***	0.47	Support	Support
H12	Home fitness intention	->	Residents' satisfaction	0.214	0.05	Does not support	Does not support

Note: -> indicates the path influence relationship, ***indicates significant at the level of P<0.001, **indicates significant at the level of P < 0.05.

TABLE 9 Modified model variable fitting index.

indicators	Squared/df	GFI	RMSEA	RMR	CFI	NFI	NNFI
The judgment standard	<3	>0.9	<0.10	<0.05	>0.9	>0.9	>0.9
Value	2.11	0.906	0.037	0.024	0.923	0.915	0.941

potentially kind of a nuisance, with the risk of affecting the relationship of neighbors. These factors in the particular historical period are likely to lead to a poor fitness environment at home for residents. However, with the end of the epidemic, people's demand for home fitness will be released, and the purchase of sports equipment will be convenient. Thus, residents are more likely to create a relatively satisfactory home fitness environment. Although the verification results indicate that residents' home fitness environment has no significant impact on the gain of home fitness ($P = 0.047$), it still has a certain impact. So, hypothesis H6 was not deleted in the correction.

The initial verification results show that residents' intention of home fitness has a certain negative impact on their satisfaction with home fitness, that is, the stronger the intention is, the lower the satisfaction will be. However, according to the analysis of $P = 0.214$, there is no significant relationship between residents' intention of home fitness and their satisfaction with home fitness. For this reason, delete H12.

In the model modification index, it is found that there is a positive influence on residents' physical literacy and residents' perception of their home fitness environment and social support. To a certain extent, residents' physical literacy reflects the comprehensive level of sports skills, sports habits, and sports knowledge. Residents with higher sports literacy prefer to obtain their required home fitness skills from various channels, which would create a good home fitness environment, and people with higher sports literacy usually have purchased a certain amount of home fitness equipment. Therefore, this study believes that residents' physical literacy has a positive impact on residents' perception of the home fitness environment and social support. So, hypotheses H13

and H14 are added. Figure 2 shows the modified model of residents' satisfaction with home fitness. In summary, one hypothesis H12 is removed and two hypotheses H13 and H14 are added.

After revision and re-examination, the model fitting degree index meets the standard, which indicates that the model is valid and has the significance of practical analysis. Finally, the satisfaction model of residents' home fitness is shown in Figure 3 (Table 9).

Table 10 shows the path relationships of each factor. The results show that hypotheses H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H13, and H14 are valid, but H6 is not supported. Since H6 has been explained, H12 is finally deleted and H13 and H14 are added.

Discussion

This study validated the satisfaction model of home fitness with structural equations and revealed the relationships among the intention of home fitness, perceived social support, perceived environment of home fitness, a gain of home fitness, residents' physical literacy, and satisfaction with home fitness. Findings (1) During the period of the epidemic at home, residents' enthusiasm for home fitness remained unchanged, online services provided strong social support, and the home fitness environment had a large room for improvement. Some residents failed to develop regular home fitness habits. During the outbreak, the sedentary lifestyle at home and the external pressure from the outbreak made the residents realize the importance of fitness and physical and mental health. Plenty of people are actively involved in home fitness, and at all levels of government, schools, media, sports departments, and service agencies have launched home fitness programs. In addition, pictures and text, video, and broadcast ways to promote fitness at home have been achieved successfully, but residents' home fitness environment still needs to be improved (12). Residents' participation degree in home fitness is high, but the time and frequency of participation are not satisfactory, which indicates that a large number of residents do not develop the habit of home fitness. (2) The perception of social support and home fitness environment have a positive impact on residents' home fitness harvest and satisfaction with home fitness. During the epidemic period, home fitness gained unprecedented social support. Professors and doctoral supervisors went online to popularize fitness knowledge, and physical education teachers became "18-line anchors," while handsome men and women from various sports service agencies opened online live-streaming fitness courses. People are creative and use everyday objects to create a fitness environment at home. For example, rice bags can be used instead of kettlebells, balloons can be used as air volleys, and tables can be used as table tennis tables (13). The more social support residents get, the better the home fitness environment will be; the more residents gain from home fitness, the higher their satisfaction with home fitness will be. (3) Residents' physical literacy directly affects residents' intention to exercise at home, perception of the environment of exercise at home, perception of social support, the harvest of exercise at home, and satisfaction with exercise at home (14). To a great extent, the level of physical literacy determines the size of a person's scientific fitness ability. The study found that physical literacy had a positive impact on the intention to exercise at home, the perception of social support, the perception of the environment of exercise at home, the harvest of exercise at home, and the satisfaction with exercise at home. The higher the sports quality, the more they can understand the benefits of physical fitness; the more people gain the benefits of physical fitness, the higher the loyalty to physical exercise will be; and good sports attainments can help to have a better channel to gain fitness knowledge, choose to science for their own home, and know how to make and use their living environment to create a satisfied and comfortable fitness environment at home with satisfaction.

According to the results of this study, during the outbreak of residents at home, the satisfaction degree of fitness is not high, especially the social support score with home fitness environment is not very outstanding. The major cause is that, on the one hand, physical sports consumption has been affected by the epidemic, which to a certain extent causes some residents to be unsatisfied with the exercise environment at home. However, after the end of the epidemic, physical sports consumption is expected to peak, especially home fitness equipment consumption. Persistence is the most important thing in sports and fitness. Although more people participated in the in-home exercise, the time and frequency of the exercise do not meet the expectations. First, the communication and guidance methods of home exercise content have been affected to some extent in the epidemic, and home exercise can only be promoted through offline channels such as the Internet and TV. Second, the content and form of home fitness are limited to a certain extent, which makes it difficult for some providers of home fitness content to find suitable shooting sites and filming equipment, and effectively organize filming personnel and logistics support personnel. Moreover, with the fierce outbreak of the epidemic, the time for shooting, organizing, and planning is relatively limited. Finally, such a large fitness activity at home is the first time to appear, so it takes time to get into the habit of home fitness. And scenes, space, during the outbreak of the above factors are affected by a certain, so residents fitness time and frequency are not high also understandable. It is believed that after the outbreak, with the recovery of the social order and continuous improvement of the high-quality fitness services, content at home and satisfaction with fitness at home will increase.

The emergence of the epidemic has forced departments at all levels, schools, sports enterprises, sports institutions, and others to think about solutions to home fitness. The importance

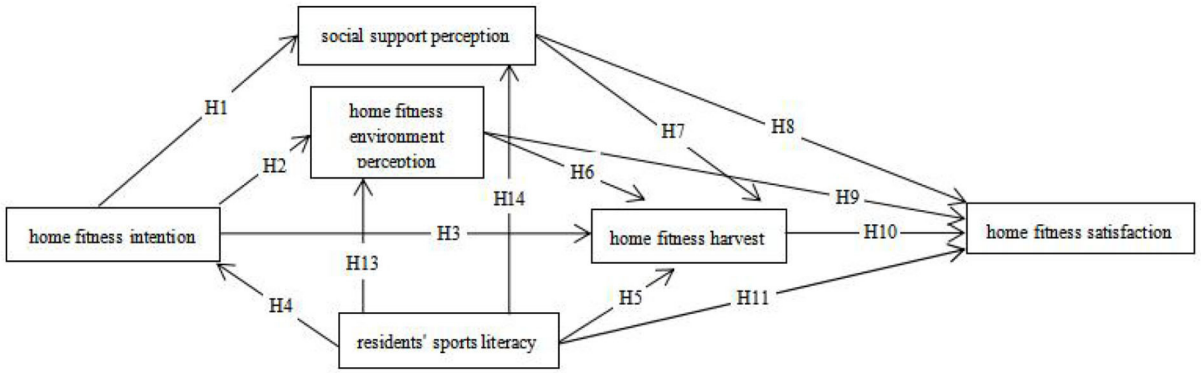


FIGURE 2
Modified model of residents' satisfaction with home fitness.

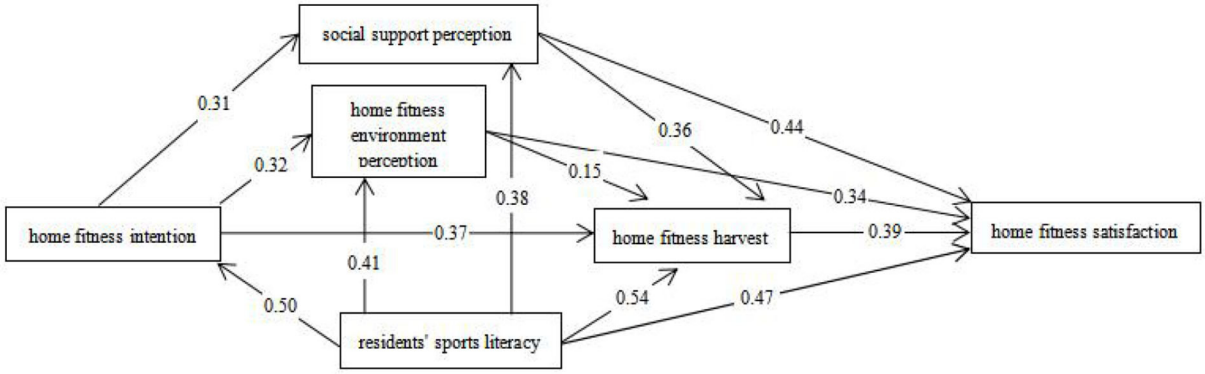


FIGURE 3
Satisfaction model of residents' home fitness.

of home fitness has received unprecedented attention from the whole society. Affected by various isolations and restrictions during the epidemic, the emerging intelligent fitness scenes and equipment have attracted much attention. In the future, home-based fitness will be characterized by diversification, intelligence, individualization, and entertainment. Internet, big data, 5G, and cloud technology will play an important role in the change. The outbreak of the COVID-19 epidemic is not only a heavy blow for the sports fitness service industry and equipment manufacturing industry but also an opportunity for reform. The outbreak will deeply affect people's lifestyles. Sports service departments should speed up the use of innovative technologies to build a new national fitness service platform and promote the upgrading and improvement of national fitness services.

Sports service departments should innovate their approaches to sports service, adopt the combination of online and offline approaches, and vigorously promote the establishment of digital and intelligent home fitness service mechanisms. Change the way of work; gradually change from offline to offline and online combination; gradually expand the path of sports and fitness services based on grassroots, community, family, and cooperation; deepen technological innovation; further, increase the guarantee of home fitness service funds; and build and constantly improve the home fitness service mechanism.

Conclusion

The sudden epidemic has caused great trauma to the mental health of people, especially those in the worst-hit areas and medical workers. In this period, sports have played a unique and irreplaceable role in regulating the psychological health of people. Electronic publications such as home fitness manuals have a calming and soothing effect. The square cabin hospital, transformed from the stadium, also played a significant role in the fight against the epidemic. After the end of the epidemic, residents' sports consumption will usher in a new upsurge, and

TABLE 10 Test of significance of path coefficient of model structure variables.

Assuming that	Path to the relationship between			P	Path coefficient	Conclusion 1 $P < 0.05$	Conclusion 2 $P < 0.001$)
H1	Home fitness intention	->	Perception of social support	***	0.31	Support	Support
H2	Home fitness intention	->	Home fitness environment perception	***	0.32	Support	Support
H3	Home fitness intention	->	Residents harvest	***	0.37	Support	Support
H4	Residents' physical literacy	->	Home fitness intention	***	0.50	Support	Support
H5	Residents' physical literacy	->	Residents harvest	***	0.54	Support	Support
H6	Home fitness environment perception	->	Residents harvest	0.132	0.15	Support	Does not support
H7	Perception of social support	->	Residents harvest	***	0.36	Support	Support
H8	Perception of social support	->	Residents' satisfaction	***	0.44	Support	Support
H9	Home fitness environment perception	->	Residents' satisfaction	***	0.34	Support	Support
H10	Residents harvest	->	Residents' satisfaction	***	0.39	Support	Support
H11	Residents' physical literacy	->	Residents' satisfaction	***	0.47	Support	Support
H13	Residents' physical literacy	->	Home fitness environment perception	***	0.41	Support	Support
H14	Residents' physical literacy	->	Perception of social support	***	0.38	Support	Support

Note: -> indicates the path influence relationship, ***indicates significant at the level of $P < 0.001$, **indicates significant at the level of $P < 0.05$.

the upstream and downstream of the sports industry chain will absorb a large number of workers, which has played a positive role in social stability. At present, it is a critical period for China to realize sports power and build a healthier China. The outbreak makes people and the government realize the importance of physical fitness, which provides opportunities to discover and meet the needs of residents' sports fitness. The outbreak is also a good time to improve the public's sports attainments. Home exercise cannot rely only on the spontaneous behavior of the individual. Each department should better seize the opportunity to vigorously promote the national fitness and healthy life concept. At the same time, scientific knowledge of fitness should be publicized and actively encourage residents to form the habit of lifelong exercise, thus creating a fitness atmosphere as well as comprehensively improving the sports literacy of people.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DC and BZ jointly conducted a statistical analysis on the model of this article and conducted statistics together. BZ completed the frame writing of the entire article. DC supplemented the data, arranged the data, and polished the article. Both authors contributed to the article and approved the submitted version.

Funding

This research study was sponsored by the General Project of 2020 Zhejiang Philosophy and Social Science Planning: Thrift Management Model of Large-Scale Sports Events and enlightenment to Hangzhou Asian Games (Project No. 20NDJC226YB); The Fundamental Research Funds of Shandong University (Project No. 2020GN029); Shandong Social Science

Planning and Research Project (Project No. 21DTYJ02); 2022 Shaoxing Higher Education Teaching Reform Research Project; and 2021 Zhejiang College of Construction Technology Innovation Team: Asian Olympic Culture Research Team (No. 2021.45.9B).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Woon FAC, Omkar D, Aaker JL, Aboltins K, Rivza B, Adusei C, et al. Effectiveness of loyalty programs in customer retention: a multiple mediation analysis. *Jindal J Bus Res*. (2021) 10:7–32. doi: 10.1177/22786821211000182

2. Fei D, Xiaoxi W, Xiaonan L. A model for the satisfaction of preschool students' education practice based on structural equation. *Preschool Educ Res*. (2018) 2018:36–45. doi: 10.13861/j.cnki.sece.2018.10.004

3. Haiyan H, Yikun K. Influence of the fit between sports events and city image on audience satisfaction and revisit intention. *China Sports Sci Technol*. (2018) 54:12–20. doi: 10.16470/j.csst.201804002

4. Zhili C, Mina L, Xuebing L, Lei Z, Wei Y, Chengqian X. Research on public sports resources, sports participation and public sports service satisfaction. *Sports Sci*. (2016) 37:77–87.

5. Saez I, Solabarrieta J, Rubio I. Motivation for physical activity in university students and its applications relation with gender, amount of activities, and sport satisfaction. *Sustainability*. (2021) 6:3183. doi: 10.3390/su13063183

6. Li X, Che L, Yuan J, Guo S, Wang K. Physical exercise activities, medical consumption and health satisfaction: a survey based on the sports participation of Beijing-Tianjin-Hebei Urban Residents. *J Wuhan Inst Phys Educ*. (2019) 53:34–42. doi: 10.15930/j.cnki.wtxb.2019.07.005

7. Zhang Y. Study on the sociological constraints of urban residents' exercise and fitness behavior. *J Beijing Sports Univ*. (2019) 42:85–94. doi: 10.19582/j.cnki.11-3785/g8.2019.07.010

8. Wang X. Frontier hotspots, topic clustering and expansion space of international sports literacy research. *J Beijing Sport Univ*. (2019) 42:102–16. doi: 10.19582/j.cnki.11-3785/g8.2019.10.012

9. Sheng PN, Xiang CP, Yun CP, Wen WZ. Home fitness and rehabilitation support system implemented by combining deep images and machine learning using unity game engine. Sensor Mater. (2022) 34:1971–90. Available online at: https://sensors.myu-group.co.jp/sm_pdf/SM2947.pdf

10. Arne JP. Hidden in plain sight: the spatial and industrial logics of home fitness technologies. New Rev Film Television Stud. (2021) 19:485–509. doi: 10.1080/17400309.2021.1960099

11. Hung CJ, Chang HM, Lin CJ, Yen TF. Modelling the indicators of environmental risk for fitness activities. *Asian J Educ Soc Stud*. (2020). doi: 10.9734/ajess/2020/v6i130167

12. Nyenhuis SM, Greiwe J, Zeiger JS, Nanda A, Cooke A. Exercise and fitness in the age of social distancing during the COVID-19 pandemic. *J Allergy Clin Immunol Pract*. (2020) 8:2152–5. doi: 10.1016/j.jaip.2020.04.039

13. Furjan Mandic G, Bilbija B, Radas J, Ivkovic G. Impact of home fitness program on anthropological characteristics of physically active and physically inactive people. *Sport Mont*. (2018). doi: 10.26773/smj.180207

14. Zhu Q, Sun J. Evaluation on home fitness and community sports activities based on network survey. *Int J Smart Home*. (2016) 16:33-6. doi: 10.14257/ijsh.2016.10.5.29

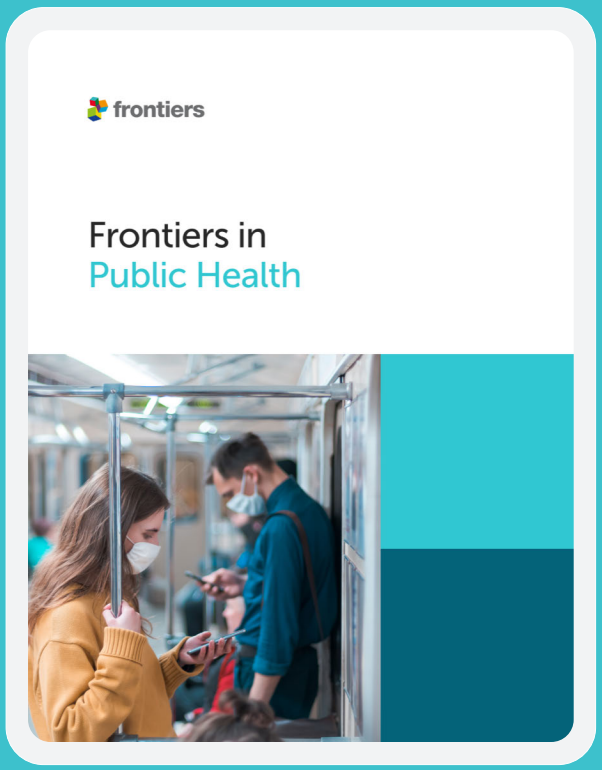
Frontiers in Public Health

Explores and addresses today's fast-moving healthcare challenges

One of the most cited journals in its field, which promotes discussion around inter-sectoral public health challenges spanning health promotion to climate change, transportation, environmental change and even species diversity.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact